

Copyright (C) 1993 - 2000 Compugen Ltd.	Gencore version 4.5		
OM protein - protein search, using sw model			
Run on: April 4, 2001, 13:03:57 ; Search time 18 Seconds	(without alignments)		
Perfect score: 1857	670.578 Million cell updates/sec		
Sequence: 1 MTGPKHKCECKSHYVGDSL.....ALAAYSTFRINRKTIGFXHF 353			
Scoring table: BLOSUM62			
Gapop 10.0 , Gapext 0.5			
Searched: 268485 seqs, 34193795 residues			
Total number of hits satisfying chosen parameters:	268485		
Minimum DB seq length: 0			
Maximum DB seq length: 2000000000			
Post-processing: Minimum Match 0%	Maximum Match 100%		
Database : A_Geneseq_36:*	Listing first 45 summaries		
1: /SIDS1/gcldata/geneseq/geneseq/AA1980.DAT:*			
2: /SIDS1/gcldata/geneseq/geneseq/AA1981.DAT:*			
3: /SIDS1/gcldata/geneseq/geneseq/AA1982.DAT:*			
4: /SIDS1/gcldata/geneseq/geneseq/AA1983.DAT:*			
5: /SIDS1/gcldata/geneseq/geneseq/AA1984.DAT:*			
6: /SIDS1/gcldata/geneseq/geneseq/AA1985.DAT:*			
7: /SIDS1/gcldata/geneseq/geneseq/AA1986.DAT:*			
8: /SIDS1/gcldata/geneseq/geneseq/AA1987.DAT:*			
9: /SIDS1/gcldata/geneseq/geneseq/AA1988.DAT:*			
10: /SIDS1/gcldata/geneseq/geneseq/AA1989.DAT:*			
11: /SIDS1/gcldata/geneseq/geneseq/AA1990.DAT:*			
12: /SIDS1/gcldata/geneseq/geneseq/AA1991.DAT:*			
13: /SIDS1/gcldata/geneseq/geneseq/AA1992.DAT:*			
14: /SIDS1/gcldata/geneseq/geneseq/AA1993.DAT:*			
15: /SIDS1/gcldata/geneseq/geneseq/AA1994.DAT:*			
16: /SIDS1/gcldata/geneseq/geneseq/AA1995.DAT:*			
17: /SIDS1/gcldata/geneseq/geneseq/AA1996.DAT:*			
18: /SIDS1/gcldata/geneseq/geneseq/AA1997.DAT:*			
19: /SIDS1/gcldata/geneseq/geneseq/AA1998.DAT:*			
20: /SIDS1/gcldata/geneseq/geneseq/AA1999.DAT:*			
21: /SIDS1/gcldata/geneseq/geneseq/AA2000.DAT:*			
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.5	8.8	354 21 Y57081
9	163.5	8.8	354 21 Y57084
10	157.5	8.5	125 15 R46627
11	156.5	8.4	252 19 W74523
12	155.5	8.4	355 16 R77034
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	262	14.1	277 13 R26044
2	262	14.1	277 18 W13654
3	262	14.1	277 20 W84087
4	187.5	10.1	360 20 Y13381
5	169.5	9.1	339 21 Y57083
6	165.5	8.9	354 21 Y57080
7	165.5	8.9	355 21 Y57082
8	163.		

PR	17-OCT-1997;	97US-0062287.
PR	21-OCT-1997;	97US-0063486.
PR	24-OCT-1997;	97US-0062814.
PR	24-OCT-1997;	97US-0062816.
PR	24-OCT-1997;	97US-0063045.
PR	24-OCT-1997;	97US-0063105.
PR	24-OCT-1997;	97US-0063121.
PR	24-OCT-1997;	97US-0063127.
PR	24-OCT-1997;	97US-0063128.
PR	27-OCT-1997;	97US-0063329.
PR	27-OCT-1997;	97US-0063347.
PR	27-OCT-1997;	97US-0063351.
PR	28-OCT-1997;	97US-0063542.
PR	28-OCT-1997;	97US-0063544.
PR	28-OCT-1997;	97US-0063549.
PR	28-OCT-1997;	97US-0063550.
PR	28-OCT-1997;	97US-0063554.
PR	29-OCT-1997;	97US-0063435.
PR	29-OCT-1997;	97US-0063704.
PR	29-OCT-1997;	97US-0063732.
PR	29-OCT-1997;	97US-0063738.
PR	29-OCT-1997;	97US-0063734.
PR	29-OCT-1997;	97US-0064215.
PR	29-OCT-1997;	97US-0063735.
PR	31-OCT-1997;	97US-0063870.
PR	31-OCT-1997;	97US-0064103.
PR	03-NOV-1997;	97US-0064809.
PR	07-NOV-1997;	97US-0064812.
PR	12-NOV-1997;	97US-0065186.
PR	17-NOV-1997;	97US-0065633.
PR	18-NOV-1997;	97US-0065634.
PR	21-NOV-1997;	97US-0066120.
PR	24-NOV-1997;	97US-0066644.
PR	24-NOV-1997;	97US-0066772.
PR	24-NOV-1997;	97US-0066666.
PR	24-NOV-1997;	97US-0066770.
PR	24-NOV-1997;	97US-0066651.
PR	24-NOV-1997;	97US-0066453.
XX	(GETH) GENENTECH INC.	
PA	Chen J,	Goddard A,
PA	Gurney J	
XX	WPI: 1999-229533/19.	
XX	DR X52252.	
XX	New isolated human genes and	
XX	gastrointestinal ulceration	
XX	Claim 12: Fig 78; 320pp; Eng	
XX	WPI: 1999-229533/19.	
XX	DR X52252.	
XX	PT	
XX	PT	
XX	PT	
XX	PS	
XX	PS	
CC	Y13344-403 represent secreted	
CC	the cDNA sequences are obtained	
CC	fetal lung, fetal kidney, fetal	
CC	known polypeptides have	
CC	associated with the presence	
CC	mucosa and the repair of acu-	
CC	ulceration and congenital mi-	
CC	with abnormal keratinocyte d-	
CC	cancers such as lung squamous	
CC	cell growth, e.g. PRO-	
CC	survival of nerve cells in	
CC	disease, ALS, neuropathies of	
CC	fibromodulin, e.g. for reduc-	
CC	as a target for anti-tumor dr-	
CC	of Usher Syndrome or Atrophic	
CC	anti-thrombotic agent; PRO28	
CC	therapeutic applications in	
CC	be used for treating problems in	
CC	vessels, or related tissue, e	
XX		

NMTAs are effective in regulating undesirable immune responses even when target determinants used as agents promoting tolerance agents have failed to induce an effective regulatory T cell response. NMTAs as agents promoting tolerance are anticipated to be safer than use of target determinants.

CC participating in the immune response. The NTTA are capable of recognition CC by substantial populations of uncommitted T cells which can be primed, or CC biased, towards regulatory responses to provide effective treatment. The CC NTTA are effective in regulating undesirable immune responses even when CC target determinants used as agents promoting tolerance agents have failed CC to induce an effective regulatory T cell response. NTTAs as agents CC promoting tolerance are anticipated to be safer than use of target CC determinants.

11 - NOV -1999 .
07 - MAY -1999 ; 99WO-US10250 .
07 - MAY -1998 ; 98US-0084636 .
(REGC) UNIV CALIFORNIA .

WPI; 2000-052905/04.

Administration of neglected target tissue antigens to modulate immune responses

Amino acid sequences Y57063-Y57091 are examples of neglected target tissue antigens. NTRAs are antigens (whole antigens or fragments) not involved in autoimmunity. These peptides and proteins are used in the method of the invention which involves administering an NTRA as an antigen based immunotherapeutic agent, to a host afflicted with an autoimmune response associated with an autoimmune disease. The immunotherapeutic agent is used to treat autoimmune diseases such as insulin dependent diabetes mellitus, multiple sclerosis, autoimmune thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory immune responses. The NTRA induces regulatory tolerance by elicitation of regulatory T cells among T cells recognizing the NTRA but not

thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory immune responses. The NNTA induces regulatory tolerance by elicitation of regulatory T cells among T cells recognizing the NNTA but not participating in the immune response. The NNTA are capable of recognition by substantial populations of uncommitted T cells which can be primed, or biased, towards regulatory responses to provide effective treatment. The NNTA are effective in regulating undesirable immune responses even when target determinants used as agents promoting tolerance agents have failed to induce an effective regulatory T cell response. NNTAs as agents promoting tolerance are anticipated to be safer than use of target determinants.

XX Sequence 354 AA;

Query Match 8.8%; Score 163.5; DB 21; Length 354;
Best Local Similarity 30.1%; Prd. No. 6.1e-09;
Matches 44; Conservative 21; Mismatches 5; Indels 29; Gaps 4;

Qy 5 GKHCKECKSHYVGDLNCEPEQLPDIRCLQDNGOQADAKCYDLHFDQTGVFHLRSPL 64
Db 135 grykce----iegl-----eddtvvald1qgv--vfpypfrl 167

Qy 65 GQYKLTFDKAREACANEATMATYNOQSYKAKYHLCAGMLETGRVAYPTAFASQNCG 124
Db 168 grylnfnheadqqacldqdaiviasfdqlydawrgldwcnagwlsd9svqyptkprecg 227

Qy 125 --SGVVGIVDYGPRPNKSEMWDVFY 148
Db 228 qgntvpgvnygfwdkdsrydvcf 253

RESULT 9
Y57084 standard; protein; 354 AA.
XX ID Y57084
AC Y57084;
XX DT 28-FEB-2000 (first entry)
DE Human proteoglycan link protein precursor #2.

XX Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response; multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis; uveoretinitis; inflammatory response.
KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;
KW autoimmune; insulin dependent; diabetes mellitus;
KW autoimmune; cellular; humoural; carbohydrate;
KW hypersensitivity; trauma; neuronal development; cell transport; enzyme;
KW infection; diagnosis; lectin; versican; aggrecan; gelosolin; saccharide;
KW receptor; cell recognition; membrane cytoplasmic protein; nucleoside.
XX OS Homo sapiens.

XX PD 11-NOV-1999.
XX PT 07-MAY-1999; 99W0-US10250.
XX PR 07-MAY-1998; 98US-0084636.
XX PA (RESC) UNIV CALIFORNIA.
XX KAUFMAN DL, Tian J, Olcott A;
XX DR WPI; 2000-052905/04.

XX PT Administration of neglected target tissue antigens to modulate immune responses -
XX Disclosure; Page 28; 79pp; English.
PS CC Amino acid sequences Y57063-Y57091 are examples of neglected target tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments) not involved in autoimmunity. These peptides and proteins are used in the method of the invention which involves administering an NNTA as an antigen based immunotherapeutic agent, to a host afflicted with an

autoimmune response associated with an autoimmune disease. The immunotherapeutic agent is used to treat autoimmune diseases such as insulin dependent diabetes mellitus, multiple sclerosis, autoimmune thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory immune responses. The NNTA induces regulatory tolerance by elicitation of regulatory T cells among T cells recognizing the NNTA but not participating in the immune response. The NNTA are capable of recognition by substantial populations of uncommitted T cells which can be primed, or biased, towards regulatory responses to provide effective treatment. The NNTA are effective in regulating undesirable immune responses even when target determinants used as agents promoting tolerance agents have failed to induce an effective regulatory T cell response. NNTAs as agents promoting tolerance are anticipated to be safer than use of target determinants.

XX Sequence 354 AA;

Query Match 8.8%; Score 163.5; DB 21; Length 354;
Best Local Similarity 30.1%; Prd. No. 6.1e-09;
Matches 44; Conservative 21; Mismatches 52; Indels 29; Gaps 4;

Qy 5 GKHCKECKSHYVGDLNCEPEQLPDIRCLQDNGOQADAKCYDLHFDQTGVFHLRSPL 64
Db 135 grykce----iegl-----eddtvvald1qgv--vfpypfrl 167

Qy 65 GQYKLTFDKAREACANEATMATYNOQSYKAKYHLCAGMLETGRVAYPTAFASQNCG 124
Db 168 grylnfnheadqqacldqdaiviasfdqlydawrgldwcnagwlsd9svqyptkprecg 227

Qy 125 --SGVVGIVDYGPRPNKSEMWDVFY 148
Db 228 qgntvpgvnygfwdkdsrydvcf 253

RESULT 10
R46627

ID R46627 standard; Protein; 1257 AA.

XX AC R46627;

XX DT 11-AUG-1994 (first entry)

XX DE Neurocan core protein.

XX KW Neurocan; cell adhesion; leukocyte; endothelial cell recognition; lipid; tissue related inflammation; allergy; cellular; humoural; carbohydrate; hypersensitivity; trauma; neuronal development; cell transport; enzyme; infection; diagnosis; lectin; versican; aggrecan; gelosolin; saccharide; receptor; cell recognition; membrane cytoplasmic protein; nucleoside.
XX OS Homo sapiens.

XX FH Key Location/Qualifiers
FT Peptide 45..55
FT /note= "Used for primer design"
FT Modified-site 220 /note= "O-linked glycosylation site"
FT Modified-site 319 /note= "O-linked glycosylation site"
FT Modified-site 339 /note= "N-linked glycosylation site"
FT Modified-site 361 /note= "N-linked glycosylation site"
FT Peptide 364..367
FT /note= "RGDS Peptide"
FT Binding-site 372..373 /note= "Chondroitin sulphate attachment site"
FT Binding-site 400 /note= "O-linked glycosylation site"
FT Binding-site 410..411 /note= "Chondroitin sulphate attachment site"
FT Modified-site 431 /note= "Chondroitin sulphate attachment site"

FT	Modified-site	/note= "O-linked glycosylation site"	PT	treatment or research of hypersensitivity and allergic diseases
FT	Modified-site	436 /note= "O-linked glycosylation site"	XX	Claim 1; Page 69-75; 105pp; English.
FT	Modified-site	445 /note= "O-linked glycosylation site"	PS	
FT	Modified-site	454 /note= "O-linked glycosylation site"	XX	
FT	Modified-site	456 /note= "O-linked glycosylation site"	CC	This sequence represents a neurocan polypeptide. This protein has
FT	Modified-site	511 /note= "O-linked glycosylation site"	CC	several biological activities, including cell adhesion, leukocyte-
FT	Modified-site	512 /note= "O-linked glycosylation site"	CC	endothelial cell recognition, tissue-related inflammation, trauma, neuronal
FT	Modified-site	572 /note= "O-linked glycosylation site"	CC	cellular and/or humoral hypersensitivity, trauma, neuronal
FT	Modified-site	576 /note= "O-linked glycosylation site"	CC	development, and cell transport and/or infection. Compositions
FT	Modified-site	579 /note= "O-linked glycosylation site"	CC	containing them can be used as modulators of these conditions, and
FT	Modified-site	655 /note= "O-linked glycosylation site"	CC	may be used as therapeutic, diagnostic, and/or research tools.
FT	Modified-site	689 /note= "O-linked glycosylation site"	CC	Neurocan peptides can be used to mimic proteins, such as lectins,
FT	Modified-site	737 /note= "O-linked glycosylation site"	CC	cell adhesion molecules, versicans, aggrecans or gelosins, as
FT	Modified-site	764 /note= "N-linked glycosylation site"	CC	receptor or effector subtypes. The protein can be used to treat
FT	Modified-site	793 /note= "O-linked glycosylation site"	CC	diseases involving a qualitative or quantitative pathological
FT	Modified-site	807 /note= "O-linked glycosylation site"	CC	abnormality of cell adhesion or leukocyte-endothelial cell recognition,
FT	Modified-site	829 /note= "O-linked glycosylation site"	CC	or a functionally associated molecule such as a membrane cytoplasmic
FT	Modified-site	857 /note= "O-linked glycosylation site"	CC	protein, lipid, carbohydrate, nucleoside, enzyme or ion.
FT	Modified-site	859 /note= "O-linked glycosylation site"	XX	
FT	Modified-site	861 /note= "O-linked glycosylation site"	Sequence	Sequence 1257 AA;
FT	Modified-site	863 /note= "O-linked glycosylation site"	Query	8 5%;
FT	Modified-site	866 /note= "O-linked glycosylation site"	Match	Score 157 5;
FT	Modified-site	867 /note= "O-linked glycosylation site"	Best Local	DB 15;
FT	Modified-site	870 /note= "O-linked glycosylation site"	Similarity	Length 1257;
FT	Modified-site	913 /note= "O-linked glycosylation site"	Matches	34.28;
FT	Modified-site	933..942 /note= "O-linked glycosylation site"	41;	pred. No. 1.8e-07;
Peptide		944..945 /note= "Used for primer design"	Conservative	Mismatches 53;
FT	Binding-site	950 /note= "Chondroitin sulphate attachment site"	Indels	Indels 7;
FT	Modified-site	967 /note= "N-linked glycosylation site"	Gaps	Gaps 3;
FT	Peptide	1136..1147 /note= "Used in primer design"		
FT	Modified-site	1164 /note= "N-linked glycosylation site"		
XX	W09403601-A.			
PN	17-FEB-1994.			
XX	03-AUG-1993;	93WO-US07306.		
XX	03-AUG-1992;	92US-0922911.		
PA	(UYY) UNIV NEW YORK STATE.			
XX	PI Margolis RK, Margolis RU,	Rauch U;		
XX	DR N PSDR; Q57710.			
XX	DR Eukaryotic neurocan polypeptide(s) with epidermal growth factor,			
PT	lectin or complement binding activity - used in the diagnosis,			

PR 22-FEB-1997; 97EP-0102951.
 XX
 PA (LANSING M.
 PA (SCHMIDT G.
 PA (UHDE/ UHLENKUERKEN J.
 XX
 PI Lansing M;
 XX
 DR WPI; 1998-449114/39.
 DR N-PSDB; V53715.
 XX
 PT Production of homogeneous polysaccharides from heterogeneous
 PT polysaccharides used for diagnosis and therapy of diseases -
 PT comprises immobilisation on support e.g. polymer matrix and
 PT selective cleavage with e.g. glucosidase or hydrolase
 XX
 PS Example 1; Fig 3; 19pp; English.
 XX
 CC This is the amino acid sequence of the human aggrecan G1-B domain,
 CC used in the method of the invention which involves the production
 CC of homogeneous polysaccharides from heterogeneous polysaccharides
 CC The oligonucleotides and polysaccharides are useful for the diagnosis
 CC and therapy of diseases .
 XX
 SQ Sequence 116 AA;
 Query Match 8.1%; Score 151; DB 19; Length 116;
 Best Local Similarity 35.8%; Pred. No. 2.e-08;
 Matches 34; Conservative 15; Mismatches 42; Indels 4; Gaps 2;
 Qy 57 VFHLRSPLCQQKLTEDKAREACANEATMATYNQLSLSYXQKAKYHILCSAGWLTGRVAYPT 116
 Db 14 vfhraistrtyidfrgracqmsaiatpeqlqayedfhqcdagwladqtvrypi 73
 Qy 117 AFASQNC---GSGGVYIVDYGPRPNKSEMMWDYFCY 148
 Db 74 htpregcyydkdefpgvrtgir-dtnetydycf 107

Search completed: April 4, 2001, 13:05:01
 Job time: 64 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:01 ; Search time 27.49 Seconds
(without alignments)
1505.071 Million cell updates/sec

Title: US-09-466-778-11
Sequence: 1 MTGPWKHCKECKSHYVGDL.....ALAAYSYFRINRKRTIGFXHF 353

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rabbit:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

Database : SPTREMBL_15:*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rabbit:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query	Match	Length	DB	ID	Description
1	17.57	94.6	897	4	Q9NRY3	Q9NRY3	09ny7 homo sapien
2	14.68	79.1	1069	4	Q8UF98	Q9uf98	09uf98 homo sapien
3	6.62	35.6	2212	4	Q93072	Q93072	093072 homo sapien
4	6.45	34.8	2570	4	Q9R115	Q9r115	09r115 homo sapien
5	2.64	14.2	275	11	Q08859	Q08859	008859 mus musculus
6	1.77	9.5	2109	13	P79787	P79787	P79787 gallus gallus
7	1.69.5	9.1	355	11	Q921X7	Q921X7	0921x7 mus musculus
8	1.62	8.7	1321	4	Q14594	Q14594	014594 homo sapien
9	1.61.5	8.7	655	11	Q88564	Q88564	088564 rattus norvegicus
10	1.61.5	8.7	1290	13	Q9WF61	Q9wf61	09wf61 gallus gallus
11	1.60.5	8.6	2394	6	Q77610	Q77610	077610 bos taurus
12	1.53	8.2	192	6	Q2817	Q2817	02817 orctocelus
13	1.53	8.2	656	6	Q77612	Q77612	077612 bos taurus
14	1.51.5	8.2	1643	6	Q77611	Q77611	077611 bos taurus
15	1.51.5	8.2	3381	6	Q77609	Q77609	077609 bos taurus
16	1.26.5	6.8	303	103	Q9TB3	Q9tb3	09tb3 sus scrofa
17	1.20.5	6.5	103	6	Q45380	Q45380	045380 orctocelus
18	1.10	5.9	396	13	Q9W654	Q9w654	09w654 gallus gallus
19	5.9	5.9	302	4	Q9UNE4	Q9une4	09une4 homo sapien

ALIGNMENTS

RESULT	1	Q9NRY3	PRELIMINARY;	PRT;	897 AA.
ID	Q9NRY3	Q9NRY3;			
AC	Q9NRY3;				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)			
DE	CDA4-LIKE PRECURSOR FELL.				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Tao Q., Zhang W., Cao X.				
RT	"Molecular cloning and characterization of human FELL sharing homology with CD44."				
RT	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.				
RL	EMBL: AF160476; AAF83981;				
DR	DR				
SQ	SEQUENCE 89 AA; 97.85 MW; EB920AF36101388 CRC64.				

QY	241	VSMFFYNDLVNGTTLQTRGSKLLITDQDPLHPTETRCVQDGRDTLENDICASNGITHVI	300	RESULT 3	
Db	729	VSMFFYNDLVNGTTLQTRGSKLLITASQDPLQTERFVDRGAIQLWQDIAASNGITHVI	788	Q93072 ID PRELIMINARY; PRT: 2212 AA.	
QY	301	SRLXKAPPVTLXHTXLGXFXXILVTVGAVALAAYSFRINRKTIGFQHF	353	Q93072 ID PRELIMINARY; PRT: 2212 AA.	
Db	789	SRPLKAPPVTLHTGLGAGIFFAILTVGAVALAAYSFRINRKTIGFQHF	841	Q93072 ID PRELIMINARY; PRT: 2212 AA.	
RESULT 2					
	Q9UF98	PRELIMINARY; PRT: 1069 AA.			
AC	Q9UF98;				
DT	01-MAY-2000 (TREMBLrel. 13, Created)				
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)				
DE	HYPOTHETICAL 115.7 KDA PROTEIN (FRAGMENT).				
GN	DKEF2434E0321.				
OS	Homo sapiens (Human).				
OC	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
NCBI_TaxID	9606;				
OX					
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=TESTIS;				
RA	Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;				
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ/ databases.				
DR	EMBL: AL13021; CAB61358; 1; -.				
DR	HSSP: P98066; 1; TSG.				
DR	INTERPRO: IPR000538; -.				
DR	INTERPRO: IPR000538; -.				
DR	INTERPRO: IPR000561; -.				
DR	INTERPRO: IPR000782; -.				
DR	INTERPRO: IPR002049; -.				
DR	PFAM: PF00008; EGF; 8.				
DR	PFAM: PF00193; Xlink; 1.				
DR	PRINTS: PRO0011; EGF1.				
DR	PRINTS: PRO0125; LINKMODULE.				
DR	PROSITE: PS0002; EGF_1; UNKNOWN_2.				
DR	PROSITE: PS0012; EGF_1; UNKNOWN_1.				
DR	PROSITE: PS01186; EGF_2; 5.				
DR	PROSITE: PS01248; LAMININ_TYPE_EGF; 2.				
DR	KW				
FT	Hypothetical protein.				
FT	NON_TER 1 AA: 115666 MW: 04B0960932164D63 CRC64;				
SQ	SEQUENCE 1069 AA: 115666 MW: 04B0960932164D63 CRC64;				
Query Match	79.1%; Score 1468; DB 4; Length 1069;				
Best Local Similarity	80.5%; Pred. No. 1.2e-133;				
Matches	284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;				
QY	1	MTGPQKHKCECKSHYVGDLNCEPQELPTDRC1QDNGOCHADAKCVDLHFQDTTVGVFHL	60	QY 2 TGPCKHKCECKSHYVGDLNCEPQELPTDRC1QDNGOCHADAKCVDLHFQDTTVGVFHL 60	5;
Db	697	MTGPQKHKCECKSHYVGDLNCEPQELPTDRC1QDNGOCHADAKCVDLHFQDTTVGVFHL	756	Db 1796 TGLNTRCECHAGYVGDLQCLESEPPVDRCLGQPPCHSDAMCTDHFQEKRAGYFHL 1855	
QY	61	RSPLGQKLTEDKAREACANEAAATMAYNQLSYXQAKYHLCAGMLETRGVAYPTAFAS	120	QY 61 RSPLGQKLTEDKAREACANEAAATMAYNQLSYXQAKYHLCAGMLETRGVAYPTAFAS 120	
Db	757	RSPLGQKLTEDKAREACANEAAATMAYNQLSYXQAKYHLCAGMLETRGVAYPTAFAS	816	Db 1856 QATSGPVGPNFSEAAECAQGAVIASPQLSAQQLGPHCLMGWLANGSTAHPPV 1915	
QY	121	QNCGSGGVGIVDYGPRPNMSMWDYQFVTRMKGDNCTXVGYVQDFSTSGNLQVMSFP	180	QY 121 QNCGSGGVGIVDYGPRPNKSEMWDYQFVTRMKGDNCTXVGYVQDFSTSGNLQVMSFP	179
Db	817	QNCGSGGVGIVDYGPRPNKSEMWDYQFVTRMKG-----SAGLFQQLSSRP	860	Db 1916 ADCGNRGNGRGTIVSLGARKNLSERWDAYCVRQDVACRNRGFGVGDG1STNCKGKLDVLAAT 1975	
QY	181	SLTNPLTTEVLAYNSSARGRAFELHLDLSIRGTLFQPNQSGLGENETLSGRDIEHHLAN	240	QY 180 PSLTNPLTTEVLAYNSSARGRAFELHLDLSIRGTLFQPNQSGLGENETLSGRDIEHHLAN	239
Db	861	-----RTPDDLSIROTFLFVQNSGIGENETLSGRDIEHHLAN	900	Db 1976 ANFSTYGMGLGYNATQRLGDFLDFDDELTYKTFVYNEGFVDNNTLSGPDLHLAS 2035	
QY	241	VSMFFYNDLVNGTTLQTRGSKLLITDQDPLHPTETRCVQDGRDTLENDICASNGITHVI	300	QY 240 NVSMFFYNDLVNGTTLQTRGSKLLITDQDPLHPTETRCVQDGRDTLENDICASNGI 296	
Db	901	VSMFFYNDLVNGTTLQTRGSKLLITASQDPLQTERFVDRGAIQLWQDIAASNGITHVI	960	Db 2036 NATLISAN-ASQGKLLPAHSGSLLISIAGPDNNSWAPTAGTIVVSRITIWDIMATNGI 2094	
QY	301	SRLXKAPPVTLXHTXLGXFXXILVTVGAVALAAYSFRINRKTIGFQHF	353	QY 297 THVISXKAPPAPYTLXHXLGXFXXILVTVGAVALAAYSFRINRKTIGFQHF	
Db	961	SRPLKAPPVTLHTGLGAGIFFAILTVGAVALAAYSFRINRKTIGFQHF	1013	Db 2095 IHALASPLAPPQPAVLAPEAPPVAAAGVAGLYLARGKDMGF 2150	

RESULT	4						
Q9NY15		PRELIMINARY;		PRT;	2570 AA.		
ID	Q9NY15;						
AC	Q9NY15;						
DT	01-OCT-2000 (TREMBLrel. 15, Created)						
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)						
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)						
DE	STABILIN-1.						
GN	STAB1.						
OS	Homo sapiens (Human).						
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
OX							
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	Politz O., Guillot P., Gratchev A., Schledzewski K., Birk R., Hakki N., Tebbe B., Orfanius C.E., Goerdt S.,						
RT	*Stablin-1: an endothelial-macrophage member of the fasciclin domain containing protein family associated with angiogenesis.;						
RL	Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.						
DR	EMBL: AJ215213; CAB1827.1; -.						
SEQUENCE	2570 AA; 275346 MW;	3123FABD7C8E2BF8	CRC64;				
Query Match	34.8%	Score 645.5;	DB 4;	Length 2570;			
Best Local Similarity	39.3%	Pred. No. 2.4e-53;					
Matches	142;	Conservative	48;	Mismatches 154;	Indels 17;	Gaps 7;	
Qy	2 TGGPKHKCECKSHYVGDLNC-EPEQLP1DRCLODNGQCHADAKCVDLHFQDTTGVFHL	60					
Db	21153 TGLNTRRCECHAGYVGDLQCLEESEPPVDRCLGQPPCPHSARMCTDQHFOEKRAVYFHL	2212					
Qy	61 RSPLQYKJLTFDKAREACANEATMATYNQSLYQKAYHLCASAGWLTGRVAYPTAFAS	120					
Db	2213 QATSPYGLNNEAAACFAGAVLASEAAGLQFHLCLMGLWANGSTAHPPVFPV	2272					
Qy	121 QNCGSGVYGVITYDGYGRPKNSMDVECYRMKDYNCTXVGYVGDFGS-YSGNLQLVLMSF	179					
Db	2273 ADCGGGRVGVSSGARKNLSERWDAYCFRQVADCRNGFVEDGSDTICNGKLLDVLAAAT	2332					
Qy	180 PSLTNFLTEVLYAISNSSARGRAFLEHLDLSIRGLTFXPQNSQGENETLSSGRDIEHHLA	239					
Db	2333 ANFSTFYGMGLGYZANATQGLDFLDELDLTLYKTFLPVNNEFYDNTMILSSPDLLEHAS	2392					
Qy	240 NVSMFYNDLNGTTLQTRLGSKLITDR--ODPLHPTETRCDGRTLENDICASNGI	296					
Db	2393 NATLSSAN-ASQKLLPAHSGLSSLIISDAGPDNSWAVPAPGVVVSVRLIVDIMAFIGI	2451					
Qy	297 THVISRXLKAPPAPVTLX---HTXLGXGIXXILVGTAV--ALAAYSFPINRKTG	349					
Db	2452 THALASPLAAPQQAQVLAPEAPVVAAGVG--AVLAGALIGLVAGALYRARGPTG	2507					
Qy	350 F 350						
Db	2508 F 2508						
RESULT	5						
Q08859		PRELIMINARY;		PRT;	275 AA.		
ID	Q08859;						
AC	Q08859;						
DT	01-JUL-1997 (TREMBLrel. 04, Created)						
DT	01-JUL-1997 (TREMBLrel. 04, Last sequence update)						
DE	TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN TSG-6 (HYALURONATE-BINDING PROTEIN).						
GN	TNFIP6 OR TNFAIP6 OR TSG6.						
OS	Mus musculus (Mouse),						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
OX							
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=98087423; PubMed=9427551;						
RA	Fulop C., Kamath R.V., Li Y., Otto J.M., Salustri A., Olsen B.R., Giant T.T., Hascall V.C.						
RT	"Coding sequence, exon-intron structure and chromosomal localization of murine TNF-stimulated gene 6 that is specifically expressed by expanding cumulus cell-oocyte complexes."						
RT	Gene 202:95-102(1997).						
RL	EMBL; U83903; AAC53527.1; -.						
DR	HSSP; P98066; IISG.						
DR	MGI; MGI:1195266; TfniP6.						
DR	INTERPRO; IPR000358; -.						
DR	INTERPRO; IPR000859; -.						
PFAM; PF00193; Xlink; 1.							
PFAM; PF00431; CUB; 1.							
PROSITE; PS01180; CUB; 1.							
DR	PROSITE; PS01241; LINK; 1.						
DR	PRODOM; PD000918; -.						
SEQUENCE	275 AA;	30924 MW;	1CD2472282608F9	CRC64;			
Query Match	14.2%	Score 264;	DB 11;	Length 275;			
Best Local Similarity	43.3%	Pred. No. 1.2e-17;					
Matches	45;	Conservative 16; Mismatches 43;	Indels 0;	Gaps 0;			
Qy	52 DTTVGFLHRSPLQYKLTEDKAREACANEATMATYQNLSYXQKAYHLCASAGWLETGR	111					
Db	32 EQAGGYHREARAGRYKLYTQEAKAYVCFGRGLTQEAEARQVYCAAGWMAKGR	91					
RESULT	6						
ID	P79787						
AC	P79787;						
DT	01-MAY-1997 (TREMBLrel. 03, Created)						
DT	01-MAY-1997 (TREMBLrel. 03, Last sequence update)						
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)						
DE	CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN.						
OS	Galus gallus (Chicken).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Archosauvia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.						
OC							
NCBI_TAXID=2031;							
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	TISSUE-BRAIN;						
RX	MEDLINE=9626324; PubMed=8965652;						
RA	Li H., Domowicz M., Henrig A., Schwartz N.B.						
RT	"S103L reactive chondroitin sulfate proteoglycan core protein expressed in developing chick brain and cartilage is encoded by a single gene."						
RT	Brain Res. Mol. Brain Res. 36:309-321(1996).						
DR	EMBL; U78555; AAC6055.1; -.						
DR	HSSP; P08709; 1BP9.						
DR	INTERPRO; IPR000152; -.						
DR	INTERPRO; IPR000436; -.						
DR	INTERPRO; IPR000538; -.						
DR	INTERPRO; IPR000561; -.						
DR	INTERPRO; IPR000742; -.						
DR	INTERPRO; IPR001304; -.						
DR	INTERPRO; IPR001881; -.						
DR	INTERPRO; IPR003006; -.						
PFAM; PF00008; EGF; 1.							
PFAM; PF00047; Ig; 1.							
PFAM; PF00059; lectin_c; 1.							
PFAM; PF00084; sushi; 1.							
PFAM; PF00193; Xlink; 4.							
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.						
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.						
DR	PROSITE; PS000615; C_TYPELECTIN_1; 1.						

Qy	226 NETUSGR-DIEHHIANVSMFFYNDLVLNGTTLQTRGLSKLLIDPQLHPTETRCVGRD 284
Db	368 DEPSEETPEH-----DLI-AEILPELGMHLSEDED-----EECANATD 409
Qy	285 -TLEWDICASNGITHVISRXLKAFA 309
Db	410 VTTTPSVQYING-KHVVTVPKDEA 434
RESULT 11	
ID 077610	PRELIMINARY; PRT; 2394 AA.
AC 077610;	
DT 01-NOV-1998 (TREMBREL 08, Created)	
DT 01-NOV-1998 (TREMBREL 08, Last sequence update)	
DT 01-OCT-2000 (TREMBREL 15, Last annotation update)	
DE VERSICAN V1 SPlice-VARIANT PRECURSOR.	
OS Bos taurus (Bovine).	
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	
OC Bovidae; Bovinae; Bos.	
OC NCBI_TaxID=9913;	
RN [1]	
SEQUENCE FROM N.A. RX	
SCHMIDLEIN field. M., Dours-Zimmermann M.T., Winterhalter K.H., RA ZIMMERMANN D.R.; RA	
"Versican V2 is a major extracellular matrix component of the mature bovine brain." RT	
RL Biol Chem, 273:15758-15764 (1998).	
DR HSSP; P01132; 1EGF.	
DR INTERPRO; IPR000152; -.	
DR INTERPRO; IPR000436; -.	
DR INTERPRO; IPR000538; -.	
DR INTERPRO; IPR000561; -.	
DR INTERPRO; IPR000742; -.	
DR INTERPRO; IPR001304; -.	
DR INTERPRO; IPR001438; -.	
DR INTERPRO; IPR001881; -.	
DR INTERPRO; IPR003006; -.	
DR PFAM; PF00008; EGF; 2.	
DR PFAM; PF000047; ig; 1.	
DR PFAM; PF00059; lectin_c; 1.	
DR PFAM; PF00193; Xlink; 2.	
DR PRINTS; PRO010; EGF_BLOOD.	
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.	
DR PROSITE; PS00615; C_TYPELECTIN_1; 1.	
DR PROSITE; PS01186; EGF_2; 1.	
DR PROSITE; PS01187; EGF_CA; 1.	
DR PROSITE; PS01241; LINK; 2.	
DR PRODOM; PD000918; -; 2.	
KW Signal; Glycoprotein; EGF-like domain.	
FT SIGNAL 1 20 POTENTIAL.	
FT CHAIN 21 2394 VERSICAN V1 SPlice-VARIANT.	
SEQUENCE 2394 AA; 261886 MW; B82A3E10FC5BD990 CRC64;	
RESULT 12	
ID 002817	PRELIMINARY; PRT; 192 AA.
AC 002817;	
DT 01-JUL-1997 (TREMBREL 04, Created)	
DT 01-JUL-1997 (TREMBREL 04, Last sequence update)	
DT 01-OCT-2000 (TREMBREL 15, Last annotation update)	
DE LARGE AGGREGATING CARTILAGE PROTEOGLYCAN CORE PROTEIN (FRAGMENT).	
OS Oryctolagus cuniculus (Rabbit).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.	
OC NCBI_TaxID=9986;	
RN [1]	
SEQUENCE FROM N.A.	
RC TISSUE=Cartilage;	
RA Nishimura M., Noshiro M., Kawamoto T., Nakamasu K., Hamada T., RA Sabato Y.; RA Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; AB004812; BA20524.1; -.	
DR HSSP; P98066; 1TSG.	
DR INTERPRO; IPR000538; -.	
DR INTERPRO; IPR003006; -.	
DR PFAM; PF000047; ig; 1.	
DR PFAM; PF00193; Xlink; 1.	
DR PRINTS; PR01265; LNKMODULE.	
DR PROSITE; PS00220; IG_MHC; UNKNOWN_1.	
DR PROSITE; PS01241; LINK; 1.	
FT NON_TER 1 192 192	
FT NON_TER 192 192 192	
SQ SEQUENCE 192 AA; 21606 MW; 44EA35FFA92CEBB8CC CRC64;	
Query Match 8.28; Score 153; DB 6; Length 192;	
Best Local Similarity 35.8%; Pred. No. 4.5e-07;	
Matches 34; Conservative 15; Mismatches 42; Indels 4; Gaps	
Qy 57 VFHILSPPLGQKLFDKARBCANEAATATYNOQLSYXOKAKYHILCAGWLETGRVAYPT 116	
Db 97 VFHYRAISTRYTLDEQRAQCLNSAITATPEQIQLQAAVEDGFHQCDAGWLADQTVRVPI 156	
Qy 117 AFASONC---GSGVGVGIVDYGPRPNKSENMDVFCY 148	
Db 157 HTPRECCYGDKEFFGVRYGIR-DTNEYDVFVCF 190	
RESULT 13	
ID 077612	PRELIMINARY; PRT; 656 AA.
AC 077612;	
DT 01-NOV-1998 (TREMBREL 08, Created)	
DT 01-NOV-1998 (TREMBREL 08, Last sequence update)	
DT 01-OCT-2000 (TREMBREL 15, Last annotation update)	
DE VERSICAN V3, SPlice-VARIANT PRECURSOR.	
OS Bos taurus (Bovine).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	
OC NCBI_TaxID=9913;	
RN [1]	
SEQUENCE FROM N.A.	
RP MEDLINE=98288320; PubMed=9624174;	
RX Schmalfield M., Dours-Zimmermann M.T., Winterhalter K.H., Zimmermann D.R.; RX	
"Versican V2 is a major extracellular matrix component of the mature	

RT	bovine brain. ⁷	
RL	J. Biol. Chem. 273:15758-15764 (1998).	
EMBL	AF060459; AAC24361.1; -.	
DR	HSSE; P01132; IEGF.	
DR	INTERPRO; IPR00152; -.	
DR	INTERPRO; IPR00436; -.	
DR	INTERPRO; IPR00538; -.	
DR	INTERPRO; IPR00561; -.	
DR	INTERPRO; IPR00742; -.	
DR	INTERPRO; IPR01304; -.	
DR	INTERPRO; IPR01438; -.	
DR	INTERPRO; IPR01881; -.	
DR	INTERPRO; IPR0006; -.	
PFAM	PF00008; EGF; 2.	
DR	PFAM; PF00047; ig; 1.	
DR	PFAM; PF00059; lectin_C; 1.	
DR	PFAM; PF00084; sushi; 1.	
DR	PFAM; PF00193; X110k; 2.	
DR	PRINTS; PR00010; EGFBLOOD.	
DR	PROSTE; PS00010; ASX_HYDROXYL; UNKNOWN_1.	
DR	PROSTE; PS00022; EGF_1; UNKNOWN_2.	
DR	PROSTE; PS000615; C_TYPELECTIN_1; 1.	
DR	PROSTE; PS01186; EGF_2; 1.	
DR	PROSTE; PS01187; EGF CA; 1.	
DR	PROSTE; PS01241; LINK; 2.	
DR	PROSTE; PS50041; C_TYPELECTIN_2; 1.	
DR	PRODOM; PD000918; -; 2.	
KW	Signal; Glycoprotein; EGF-like domain.	
FT	SIGNAL; 1 ²⁰ POTENTIAL.	
CHAIN	21 656 AA; 74753 MW; F8FF153BD10C7AB9 CRC64;	POTENTIAL.
SEQUENCE	656 AA; 74753 MW; F8FF153BD10C7AB9 CRC64;	POTENTIAL.

Query Match	8.2%	Score 153;	DB 6;	Length 656;
Best Local Similarity	26.8%	Pred. No. 2.3e-06;		
Matches 78;	Conservative	Mismatches 103;	Indels 84;	Gaps 15;
QY	51	QDT ---T'VG- -VFHLRSPLGGQKLTFDKAREACANEATMATYQNLQSYQAKAYHLCASG 105		
Db	141	QDTIVSLLTVCQVFHYRAATSRVTFNEMACKACVYDIAVATPEQLHAAVEDGFEQCDAG 200		
QY	106	WLETRGAVAYPTAFAESQNCGS --GVVGVIVDWPYGPRENKSEWMDVPCYRMDVNCYXKGVY 162		
Db	201	WLSDOTVRFPIRVREGCYGDMKGKEVRYTGYERA -PHETYDVICY - - - - -VDHL 249		
QY	163	-GDGFPSYSGNLLQVLMSPFLTNFLTEVLAISNSAR - - - - -GRAFLEHL 206		
Db	250	DGDVFH - - - - -ITAPKFTPEAGEECKTODARLATVGEIQAQAWRNGFDRCDYGWL 300		
QY	207	TDLSTRG - -T'LEXPQNSG - -LG - - - - -ENE - - - - -TLSGRDIEHHLANWSMFFYNDLYN 251		
Db	301	LDASVRYHPVYVARAQCGGGGLGVRTYRPNQNTGFPPTPSRDAYCFKRDCKRMPCLN 360		
QY	252	GTTQIOTRGLGSKLLTIDQDPLHPIET - - -RCVDS - - -RTDLEMIDICASN 294		
Db	361	GCT - - - - -CYPETTSYVCTCVCVGFYSGDRCELDECHSN 394		
RESULT	14			
	077611	PRELIMINARY;	PRT;	1643 AA.
AC	077611;			
DT	01 - Nov - 1998	(TREMBLel.	08,	Created)
DT	01 - Nov - 1998	(TREMBLel.	08,	Last sequence update)
DT	01 - Oct - 2000	(TREMBLel.	15,	Last annotation update)
VERSICAN	V2	SPLICER-VARIANT	PRECURSER.	
OS	Bos taurus (Bovine)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
NCBI_TAXID	=9913;			
RN	SEQUENCE FROM N_A;			
RP	[1]			

DR HSSP; P01132; 1EGF;
 DR INTERPRO; IPR001052;
 DR IPR001036;
 DR INTERPRO; IPR000338;
 DR INTERPRO; IPR000561;
 DR INTERPRO; IPR000742;
 DR INTERPRO; IPR001304;
 DR INTERPRO; IPR001438;
 DR INTERPRO; IPR01881;
 DR INTERPRO; IPR003006;
 DR PFAM; PF00008; EGF; 2.
 DR PFAM; PF00047; ig; 1.
 DR PFAM; PF00059; lectin_C; 1.
 DR PFAM; PF00084; sushi; 1.
 DR PFAM; PF01192; XLINK; 2.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE; PS00002; EGF; UNKNOWN_2.
 DR PROSITE; PS00015; C_TYPELECTIN_1; 1.
 DR PROSITE; PS01186; EGF; 2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS01241; LINK; 2.
 DR PROSITE; PS50041; C_TYPELECTIN_2; 1.
 DR PRODOM; PD000918; ; 2.
 KW Signal; Glycoprotein; EGF-like domain.
 FT SIGNAL; 1 20 POTENTIAL.
 CHAIN 21 3381 VERSICAN_V0 SPLICE- VARIANT;
 SEQUENCE 3381 AA; 369987 MW; F09716FA7778D459 CRC64;

Query Match 8.2% Score 151.5; DB 6; Length 3381;
 Best Local Similarity 36.8%; Pred. No. 2.0e-05;
 Matches 39; Conservative 13; Mismatches 45; Indels 9; Gaps 4;
 Matches 39; Conservative 13; Mismatches 45; Indels 9; Gaps 4;
 Qy 51 QDT --TVG--VPHLRSPLGQYKLTDFKAREACANEATMAYNQLSYXOKARYHILCSAG 105
 Db 141 QDTVSLTVEGVVPHYRATSRTLNFEMAQKACVDTGAVIATPEQLHAYEDGFEEQCDAG 200
 Qy 106 WLETFGRVAYPTAFASQNGS --GVVGIVDYGPRPKSEMMVYFCY 148
 Db 201 WLSDQTVRYPIRVPREGYGDMNGKEGRTYGFRA -PHETYDVCY 245

Search completed: April 4, 2001, 13:06:28
 Job time: 147 sec

File saved
 by user

1-4-2

Result No.	Score	Query	Match	Length	DB	ID	Description
1	264	14.2	276	1	TSG6_RABIT	P98065	oryctolagus
2	262	14.1	277	1	TSG6_HUMAN	P98066	homo sapien
3	175	9.4	2109	1	PGCA_CHICK	P07838	gallus gallus
4	171	9.2	2124	1	PGCA_RAT	P07887	rattus norvegicus
5	170	9.2	2364	1	PGCA_BOVIN	P13608	bos taurus
6	169.5	9.1	354	1	PLK_RAT	P03994	rattus norvegicus
7	169.5	9.1	356	1	PLK_MOUSE	Q94QD5	mus musculus
8	166.5	9.0	354	1	PLK_HORSE	Q28381	equus caballus
9	165.5	8.9	354	1	PLK_BOVIN	P55252	bos taurus
10	165.5	8.9	355	1	PLK_CHICK	P07354	gallus gallus
11	165	8.9	2132	1	PGCA_MOUSE	Q61222	mus musculus
12	163.5	8.8	354	1	PLK_HUMAN	P10915	homo sapien
13	161.5	8.7	354	1	PLK_PIG	P10839	sus scrofa
14	161.5	8.7	2415	1	PGCA_HUMAN	P16112	homo sapien
15	158.5	8.5	1268	1	PGCN_MOUSE	P50666	mus musculus
16	157.5	8.5	883	1	PGCB_RAT	P55068	rattus norvegicus
17	157.5	8.5	1257	1	PGCN_RAT	P55067	rattus norvegicus
18	155.5	8.4	394	1	PGCA_RABBIT	Q28670	oryctolagus
19	155.5	8.4	912	1	PGCB_BOVIN	Q28062	bos taurus
20	154.5	8.3	862	1	PGCV_MACRONE	Q28858	macaca nemestrina
21	152.5	8.2	3358	1	PGCV_MOUSE	Q62059	mus musculus
22	152.5	8.2	3562	1	PGCV_CHICK	Q90953	gallus gallus
23	150.5	8.1	883	1	PGCB_MOUSE	P01361	mus musculus
24	150.5	8.1	3396	1	PGCV_HUMAN	P13611	homo sapien
25	147	7.9	417	1	PGCB_FELIDA	P41725	felis silvestris
26	116	6.0	362	1	CD44_CRIGR	P20944	cricetulus
27	108.5	5.8	431	1	CD44_MESEAU	Q60522	m. cd44. anti
28	106	5.7	359	1	CD44_HORSE	Q05018	equus caballus
29	105	5.7	810	1	NEU1_HUMAN	Q92832	homo sapien
30	103.5	5.6	655	1	CD44_MOUSE	P15379	mus musculus
31	100	5.4	351	1	CD44_CANFIA	Q28284	canis familiaris
32	99	5.3	2907	1	FBN2_MOUSE	Q61552	mus musculus
33	98	5.3	742	1	CD44_HUMAN	P16070	h. cd44. anti

Scoring table:	BLOSUM62	Alignments				
Searched:	88757 seqs, 32294092 residues					
Total number of hits satisfying chosen parameters:	88757					
Minimum DB seq length: 0						
Maximum DB seq length: 2000000000						
Post-processing: Minimum Match 0%						
Post-processing: Maximum Match 100%						
Database :	SwissProt_39:*					
Scoring table:	BLOSUM62					
Gapop 10.0 , Gapext 0.5						
Result No.	Score	Match	Length	DB	ID	Sequence from N. A.
1	264	14.2	276	1	TSG6_RABIT	STRAIN NEW ZEALAND WHITE;
2	262	14.1	277	1	TSG6_HUMAN	MELINE-93252803; PubMed=8098034;
3	175	9.4	2109	1	PGCA_CHICK	RA
4	171	9.2	2124	1	PGCA_RAT	RL
5	170	9.2	2364	1	PGCA_BOVIN	CC
6	169.5	9.1	354	1	PLK_RAT	CC
7	169.5	9.1	356	1	PLK_MOUSE	CC
8	166.5	9.0	354	1	PLK_HORSE	CC
9	165.5	8.9	354	1	PLK_BOVIN	CC
10	165.5	8.9	355	1	PLK_CHICK	CC
11	165	8.9	2132	1	PGCA_MOUSE	CC
12	163.5	8.8	354	1	PLK_HUMAN	CC
13	161.5	8.7	354	1	PLK_PIG	CC
14	161.5	8.7	2415	1	PGCA_HUMAN	CC
15	158.5	8.5	1268	1	PGCN_MOUSE	CC
16	157.5	8.5	883	1	PGCB_RAT	CC
17	157.5	8.5	1257	1	PGCN_RAT	CC
18	155.5	8.4	394	1	PGCA_RABBIT	CC
19	155.5	8.4	912	1	PGCB_BOVIN	CC
20	154.5	8.3	862	1	PGCV_MACRONE	DR
21	152.5	8.2	3358	1	PGCV_MOUSE	DR
22	152.5	8.2	3562	1	PGCV_CHICK	DR
23	150.5	8.1	883	1	PGCB_MOUSE	DR
24	150.5	8.1	3396	1	PGCV_HUMAN	DR
25	147	7.9	417	1	PGCB_FELIDA	DR
26	116	6.0	362	1	CD44_CRIGR	DR
27	108.5	5.8	431	1	CD44_MESEAU	DR
28	106	5.7	359	1	CD44_HORSE	DR
29	105	5.7	810	1	NEU1_HUMAN	KW
30	103.5	5.6	655	1	CD44_MOUSE	FT
31	100	5.4	351	1	CD44_CANFIA	FT
32	99	5.3	2907	1	FBN2_MOUSE	FT
33	98	5.3	742	1	CD44_HUMAN	FT

Run on:	April 4, 2001, 13:04:21 ;	Search time 11.97 Seconds				
		(without alignments)				
		952.365 Million cell updates/sec				
Title:	US-09-466-778-11					
Perfect score:	1857					
Sequence:	1 MNGPKHKCECKSHVYGVGL ALAAYSYFRINRKRTIGFXHF 353					
Scoring table:	BLOSUM62	Alignments				
Gapop 10.0 , Gapext 0.5						
Result No.	Score	Match	Length	DB	ID	SEQUENCE FROM N. A.
1	264	14.2	276	1	TSG6_RABIT	STRAIN NEW ZEALAND WHITE;
2	262	14.1	277	1	TSG6_HUMAN	MELINE-93252803; PubMed=8098034;
3	175	9.4	2109	1	PGCA_CHICK	RA
4	171	9.2	2124	1	PGCA_RAT	RL
5	170	9.2	2364	1	PGCA_BOVIN	CC
6	169.5	9.1	354	1	PLK_RAT	CC
7	169.5	9.1	356	1	PLK_MOUSE	CC
8	166.5	9.0	354	1	PLK_HORSE	CC
9	165.5	8.9	354	1	PLK_BOVIN	CC
10	165.5	8.9	355	1	PLK_CHICK	CC
11	165	8.9	2132	1	PGCA_MOUSE	CC
12	163.5	8.8	354	1	PLK_HUMAN	CC
13	161.5	8.7	354	1	PLK_PIG	CC
14	161.5	8.7	2415	1	PGCA_HUMAN	CC
15	158.5	8.5	1268	1	PGCN_MOUSE	CC
16	157.5	8.5	883	1	PGCB_RAT	CC
17	157.5	8.5	1257	1	PGCN_RAT	CC
18	155.5	8.4	394	1	PGCA_RABBIT	CC
19	155.5	8.4	912	1	PGCB_BOVIN	CC
20	154.5	8.3	862	1	PGCV_MACRONE	DR
21	152.5	8.2	3358	1	PGCV_MOUSE	DR
22	152.5	8.2	3562	1	PGCV_CHICK	DR
23	150.5	8.1	883	1	PGCB_MOUSE	DR
24	150.5	8.1	3396	1	PGCV_HUMAN	DR
25	147	7.9	417	1	PGCB_FELIDA	DR
26	116	6.0	362	1	CD44_CRIGR	DR
27	108.5	5.8	431	1	CD44_MESEAU	DR
28	106	5.7	359	1	CD44_HORSE	DR
29	105	5.7	810	1	NEU1_HUMAN	KW
30	103.5	5.6	655	1	CD44_MOUSE	FT
31	100	5.4	351	1	CD44_CANFIA	FT
32	99	5.3	2907	1	FBN2_MOUSE	FT
33	98	5.3	742	1	CD44_HUMAN	FT

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:21 ; search time 11.97 Seconds

(without alignments)

952.365 Million cell updates/sec

Perfect score: 1857

Sequence: 1 MNGPKHKCECKSHVYGVGL ALAAYSYFRINRKRTIGFXHF 353

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Result No.

Score

Match

Length

DB

ID

SEQUENCE FROM N. A.

STRAIN NEW ZEALAND WHITE;

MELINE-93252803; PubMed=8098034;

RA

RL

CC

FT CONFLICT 1988 1988 F -> S (IN REF. 6).
 SQ SEQUENCE 2109 AA; 223492 MW; 7E824FD5BA2ABDA CRC64;
 Query Match 9 4%; Score 175; DB 1; Length 2109;
 Best Local Similarity 23.4%; Pred. No. 2.4e-07;
 Matches 75; Conservative 31; Mismatches 116; Indels 98; Gaps 12;
 QY 10 ECKSHYVGDGELNCEPEQQLDRCLODNGCCHA-----DARCVDLHFDQDTTVG 56
 DB 194 QCDAGCWLADOTVRPIHLPRERCYGDKEFPGVRYGVRETDETYCYAEQMOGK ---
 QY 57 VPHLSPPLGQYKLFEDKAREACANEAAATMAYNOLSYQAKAKYKHLCSAGWLETGRVAYPT 116
 DB 251 VEYATSSP---EKEFQEADKCHSUGARATTCGELYLAWKDGMDCSAGWLAIRSVRYPI 307
 QY 117 AFASONCGSSGVGVI -VDYGPRI ---PNKSEMWDVFCYRMKDVCN C-----TXKVG-YV 162
 DB 308 SRARRNCGGNLVGVRTVLNAPONTGYPPISSRYDAICYSGDDFEALVPGLFTDEVGTFL 367
 QY 163 GDGFSY-----SGNLQVAMSPLTNFL 186
 DB 368 GSAFTIQTVYQTEVPLPRLPRNTVEEARGSIAITLEPMEITATATELYAFTVLPDL -FA 425
 QY 187 TEVLYNSNSARGRAFLEHITDL -----SIRGTLFXPQNQSLGE ----- 225
 DB 426 TSVTYVETASPREENVTRYEEWAVPEEVTTSVGTAF ---ITGMAVSVSEEAIAVTA 482
 QY 226 ---NETLSGRDIEHHLANV 242
 DB 483 TPGLESASPTIEDHLVQVT 502

RESULT 4

ID PGCA_RAT STANDARD; PRT; 2124 AA.
 AC P07897;
 AC 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
 DE PROTEIN) (CSPCP).
 GN AGC1 OR AGC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1] SEQUENCE FROM N.A.
 RX MEDLINE=88087070; PubMed=3693370;
 RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
 RT "Complete primary structure of the rat cartilage proteoglycan core
 protein deduced from cDNA clones";
 RL J. Biol. Chem. 262:17757-17767(1987).
 RN [2] REVISITON TO 698.
 RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
 RL J. Biol. Chem. 263:10040-10040(1988).
 RN [3] SEQUENCE OF 1856-2124 FROM N.A.
 RX MEDLINE=86250638; PubMed=2424893;

RA Doege K., Fernandez P., Hassell J. R., Sasaki M., Yamada Y.;
 RT "Partial cDNA sequence encoding a globular domain at the C terminus
 of the rat cartilage proteoglycan.";
 RL J. Biol. Chem. 261:8108-8111(1986).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
 MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
 IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
 HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
 REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -1- SIMILARITY: DOMAIN SECRETED; EXTRACELLULAR MATRIX (BY
 CC • SIMILARITY).
 CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMprise THE AMINO
 CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3 ,

CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
 CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
 CC THE A, B, C MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
 CC AND G3.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
 CC -1- CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSCAN PROTEOGLYCAN FAMILY.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -1- EMBL; M13518; AAA41836.1;
 DR EMBL; M13518; AAA41836.1;
 DR PIR; J03485; AAA21000.1; ALT_SEQ.
 DR PIR; A33835; A23835.
 DR PIR; A28452; A8452.
 DR HSSP; P98066; ITSG6.
 DR INTERPRO; IPR000436; -.
 DR INTERPRO; IPR000495; -.
 DR INTERPRO; IPR000538; -.
 DR PFM; PF00193; Xlink; 4.
 DR PFM; PF00055; lectin_c; 1.
 DR PFM; PF00034; Sushi; 1.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS01241; LINR; 4.
 DR PROSITE; PS00615; C_TYPELECTIN; 1.
 DR PROSITE; PS50041; C_TYPELECTIN; 2.
 KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Sushi; Sushi;
 KW Receptor; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT DOMAIN 20 2124 AGGRECAN CORE PROTEIN.
 FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1910 2036 C-TYPE LECTIN.
 FT REPEAT 2040 2098 SUSHI.
 FT DOMAIN 148 140 G1_A.
 FT DOMAIN 152 268 G1_B.
 FT DOMAIN 685 798 KS.
 FT DOMAIN 801 1226 CS 1.
 FT DOMAIN 1227 1909 CS 2.
 FT DOMAIN 1910 2124 G3.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 533 554 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 509 580 BY SIMILARITY.
 FT DISULFID 533 554 BY SIMILARITY.
 FT DISULFID 607 682 BY SIMILARITY.
 FT DISULFID 631 652 BY SIMILARITY.
 FT DISULFID 1914 1925 BY SIMILARITY.
 FT DISULFID 1942 2034 BY SIMILARITY.
 FT DISULFID 2010 2026 BY SIMILARITY.
 FT DISULFID 2041 2084 BY SIMILARITY.
 FT DISULFID 2070 2097 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC,); (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC,); (POTENTIAL).

RT synthesis of an oligonucleotide probe."
 RL FBS Lett. 176:37-42 (1984).
 [6]
 RN
 RP PARTIAL SEQUENCE.
 RX MEDLINE=87005253; PubMed=3520809;
 RA Perin J.P.; Bonnet F.; Jolles P.;
 RT "Structural relationship between link proteins and proteoglycan monomers." ;
 RL FERS Lett. 206:73-77 (1986).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -1- SURCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMprise THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE C-TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G1 AND G3.
 CC -1- DOMAIN: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-LINKED (ABOUT 40) OLIGOSACCHARIDES.
 CC -1- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN ADULT AND FETAL BOVINE PROTEOGLYCANs.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on use by non profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial purposes requires a license agreement (see <http://www.isb-sib.ch/annotation/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U7615; AAB38524; 1. ;
 DR EMBL; L07053; ; NOT_ANNOTATED_CDS.
 DR PIR; A27752; A27752.
 DR PIR; A29164; A29164.
 DR PIR; B29164; B29164.
 DR PIR; E29164; E29164.
 DR PIR; G27751; G27751.
 DR HSSP; P00740; 1.RX.
 DR INTERPRO; IPR000152; -.
 DR INTERPRO; IPR000456; -.
 DR INTERPRO; IPR000538; -.
 DR INTERPRO; IPR000561; -.
 DR INTERPRO; IPR001304; -.
 DR INTERPRO; IPR001881; -.
 DR PFAM; PF00008; EGF; 1.
 DR PFAM; PF00193; Xlink; 4.
 DR PFAM; PF00059; lectin_c; 1.
 DR PFAM; PF00084; sushi; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_-; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS01241; LINK; 4.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
 DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
 KW GLYCOPROTEIN; Cartilage; Proteoglycan; Lectin; Signal; Sushi; EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
 KW SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 2364 AGGREGCAN CORE PROTEIN.
 FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.

RESULT 6

PLK_RAT	ID	PLK_RAT	STANDARD;	PRT;	354 AA.
AC	PO394;				
DT	23-OCT-1986 (Rel. 02, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).				
GN	CRTLN.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88198139; PubMed=2452158;				
RA	Rhodes C.; Doege K.; Sasaki M.; Yamada Y.;				
RT	"Alternative splicing generates two different mRNA species for rat link protein."				
RL	J. Biol. Chem. 263:6063-6067 (1988).				
RN	[2]				
RP	SEQUENCE OF 126-354 FROM N.A.				
RX	MEDLINE=8623334; PubMed=3439153;				
RA	Doege K.; Hassel J.R.; Caterson B.; Yamada Y.;				
RT	"Link protein cDNA sequence reveals a tandemly repeated protein structure."				
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3761-3765 (1986).				

Query Match 9.2%; Score 170; DB 1; Length 2364;

Best Local Similarity 38.9%; Pred. No. 7.8e-07; Indels 4; Gaps 2;

Matches 37; Conservative 12; Mismatches 42; Standard; PRT; 354 AA.

QY 57 VEHRLRSPLGQYKLTEDKAREACANEAAATMAYNOLSYQXQAKYHLCSSAGWLETGRVAYPT 116

Db 488 VFHRPQGSSSYSLTFFEEAKQACLRPTGATASPSQLQAAYEAGYEQCDAGWLDQTVRYP 547

QY 117 AFASONC---GSGCUGGIVDYGPRPNKSEMWDVPCY 148

Db 548 VSPRTPCIVGDKDSSPGVRYGVRE-PSENYDVCY 581

SEQUENCE 2364 AA; 246359 MW; 6FB83763420C3D4C CRC64;

Query Match 9.2%; Score 170; DB 1; Length 2364;

Best Local Similarity 38.9%; Pred. No. 7.8e-07; Indels 4; Gaps 2;

Matches 37; Conservative 12; Mismatches 42; Standard; PRT; 354 AA.

QY 57 VEHRLRSPLGQYKLTEDKAREACANEAAATMAYNOLSYQXQAKYHLCSSAGWLETGRVAYPT 116

Db 488 VFHRPQGSSSYSLTFFEEAKQACLRPTGATASPSQLQAAYEAGYEQCDAGWLDQTVRYP 547

QY 117 AFASONC---GSGCUGGIVDYGPRPNKSEMWDVPCY 148

Db 548 VSPRTPCIVGDKDSSPGVRYGVRE-PSENYDVCY 581

RESULT 6

PLK_RAT	ID	PLK_RAT	STANDARD;	PRT;	354 AA.
AC	PO394;				
DT	23-OCT-1986 (Rel. 02, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).				
GN	CRTLN.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88198139; PubMed=2452158;				
RA	Rhodes C.; Doege K.; Sasaki M.; Yamada Y.;				
RT	"Alternative splicing generates two different mRNA species for rat link protein."				
RL	J. Biol. Chem. 263:6063-6067 (1988).				
RN	[2]				
RP	SEQUENCE OF 126-354 FROM N.A.				
RX	MEDLINE=8623334; PubMed=3439153;				
RA	Doege K.; Hassel J.R.; Caterson B.; Yamada Y.;				
RT	"Link protein cDNA sequence reveals a tandemly repeated protein structure."				
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3761-3765 (1986).				

Query Match 9.1%; Score 169.5; DB 1; Length 354;

Best Local Similarity 30.8%; Pred. No. 8.3e-08; Mismatches 20; Indels 29; Gaps 4;

Matches 45; Conservative 20; Standard; PRT; 354 AA.

QY 5 GKKCFCOKSHVYVGDLNCEPEQLPDIRCLQDNGQCHADAKCYLHFQDTTVGVFLRSPL 64

Db 135 GRYKCE----VIEGL-

QY 65 GQYKLTDFDKAREACANEAAATMAYNOLSYQXQAKYHLCSSAGWLETGRVAYPTAFASQNC 124

Db 168 GRYNLFNFEARQCLDQDIAVTFQLYDAWGGGLDWNAGNLSGDSVQYPTKPREFCG 227

QY 125 -SGVGVIVDGPVPNKSENWDVFY 148

Db	228	QONTVPGVNNYGFWDKDKSRYDVEFCF	253	QY	65	GOYKLTFDKARECANAEATMATYNOLSYXQAKYHLC3AGWLERTGRAYPTAFASQNGG	124
RESULT	7			Db	170	GRYNLNTHEARACLDQDAVASFQDLYDAWRGGLDWCAGWLSDGSVQPITRPREPPG	229
PLK_MOUSE		STANDARD;	PRT;	QY	125	--SGVYGLVDGPRPKNSEMDVFCY	148
ID PLK_MOUSE				Db	230	GONTVPGVNNYGFWDKDKSRYDVEFCF	255
AC Q90P5;				RESULT	8		
DT 01-OCT-2000	(Rel. 40, Created)			PLK_HORSE		STANDARD;	PRT;
DT 01-OCT-2000	(Rel. 40, Last sequence update)			ID PLK_HORSE			354 AA.
DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).				AC Q2381;			
GN CRTL1.				DT 01-NOV-1997	(Rel. 35, Created)		
OS Mus musculus (Mouse).				DT 01-NOV-1997	(Rel. 35, Last sequence update)		
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				DT 15-JUL-1998	(Rel. 36, Last annotation update)		
CC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.				DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).			
RN [1]				GN CTR1L.			
RP				OS Equus caballus (Horse).			
RC STRAIN=129/SV;				OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
RA Deak F., Mates L., Krysan K., Liu Z., Szabo P.E., Mann J.R.,				OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.			
RT "Characterization and chromosomal location of the mouse link protein				RN [1]			
RT gene (Crtl1)."				RP SEQUENCE FROM N.A.			
RL Cytogenet. Cell Genet. 87:75-79 (1999).				RC TISSUE=CARTILAGE;			
CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH				RX MEDLINE=96039874; PubMed=7574168;			
CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.				RA Dugich J., Platt D.			
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.				RT "Complete primary sequence of equine cartilage link protein deduced			
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.				RT from complementary DNA."			
CC				RL Am. J. Vet. Res. 56:959-965 (1995).			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration				CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -				CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX (BY			
CC the European Bioinformatics Institute. There are no restrictions on its				CC SIMILARITY).			
CC use by non-profit institutions as long as its content is in no way				CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.			
CC modified and this statement is not removed. Usage by and for commercial				CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.			
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/				CC			
CC or send an email to license@isb-sib.ch).				CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -				CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its				CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way				CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial				CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/				CC entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC or send an email to license@isb-sib.ch).				CC or send an email to license@isb-sib.ch).			
CC				CC			
DR EMBL; AF137278; AF24166.1;				DR EMBL; X78077; CAA54987.1;			
DR EMBL; AF137275; AF24166.1; JOINED.				DR HSSP; P98066; 1TSG.			
DR PRINTS; PR01265; LINKMODULE.				DR INTERPRO; IPR005308; -.			
DR PROSITE; PS01241; LINK; 2.				DR INTERPRO; IPR003006; -.			
DR KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;				DR PFPM; PF00193; Xlink; 2.			
DR KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.				DR PROSITE; PS01241; LINK; 2.			
DR PROPEP 1 BY SIMILARITY.				DR GLYCPROTEIN; Connective tissue; Extracellular matrix; Cartilage;			
FT CHAIN 10 9 BY SIMILARITY.				KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.			
FT DOMAIN 41 133 BY SIMILARITY.				FT PROPEP 1 BY SIMILARITY.			
FT DOMAIN 178 255 LINK 1.				FT CHAIN 16 354 PROTEOGLYCAN LINK PROTEIN.			
FT DOMAIN 276 352 LINK 2.				FT DOMAIN 54 146 IG-LIKE V-TYPE DOMAIN.			
FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .).				FT DOMAIN 176 253 LINK 1.			
FT DISULFID 63 141 BY SIMILARITY.				FT DOMAIN 274 350 LINK 2.			
FT DISULFID 183 254 BY SIMILARITY.				FT DISULFID 61 139 BY SIMILARITY.			
FT DISULFID 207 228 BY SIMILARITY.				FT DISULFID 181 252 BY SIMILARITY.			
FT DISULFID 281 351 BY SIMILARITY.				FT DISULFID 205 226 BY SIMILARITY.			
FT DISULFID 306 327 BY SIMILARITY.				FT DISULFID 279 349 BY SIMILARITY.			
FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).				FT DISULFID 304 325 BY SIMILARITY.			
SEQUENCE 356 AA; 40478 MW; 3FB4BBF30E28C9F2 CRC64;				FT CABBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).			
Query Match 9.1%				FT CABBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).			
Best Local Similarity 30.8%				SEQUENCE 354 AA; 40077 MW; D815CBB7816C7172 CRC64;			
Matches 45; Conservative 20; Mismatches 52; Indels 29; Gaps 4;				Query Match 9.0%			
QY 5 GHKCECKSHYVGDLNCEPEQLPIDRCQDNGOCHAACKVCDLHFQDQTTVGYFHLRSPL 64				Best Local Similarity 30.1%			
Db 137 GRYCE-----VIECL-----EDDTAVALELQGY--VFPYFPRL 169				Pred. No. 1.5e-07;			
QY 5 GHKCECKSHYVGDLNCEPEQLPIDRCQDNGOCHAACKVCDLHFQDQTTVGYFHLRSPL 64				Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;			

RESULT 9

PLK_BOVIN

ID PLK_BOVIN

STANDARD: PRT; 354 AA.

AC P55252;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

GN CRTL1.

OS Bos taurus (Bovine).

OC Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi; Bovidae; Bovinae; Bos.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE="CARTILAGE"; PubMed=754851;

RA MEDLINE=96043243; PubMed=459124;

RA Hering T.M., Kollar J., Huynh T.D., Sandell L.J.;

RT "Bovine chondrocyte link protein cDNA sequence: interspecies conservation of primary sequence and mRNA untranslated regions."

RL Comp. Biochem. Physiol. 112B:197-203(1995).

CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.

CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; U02292; AAC04311.1; -.

DR HSSP; P98066; 1TSG.

DR INTEPRO; IPR00538; -.

DR INTEPRO; IPR003006; -.

DR PF00193; Xlink; 2.

DR PFAM; PF00047; ig; 1.

DR PROTEPE; PS01241; LINK; 2.

KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage; Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.

FT PROTEP 1 15 BY SIMILARITY.

FT CHAIN 16 354 BY SIMILARITY.

FT DOMAIN 54 146 PROTEOGLYCAN LINK PROTEIN.

FT DOMAIN 176 253 IG-LIKE V-TYPE DOMAIN.

FT DOMAIN 274 350 LINK 1.

FT DISUFLID 61 139 LINK 2.

FT DISUFLID 181 252 BY SIMILARITY.

FT DISUFLID 205 226 BY SIMILARITY.

FT DISUFLID 279 349 BY SIMILARITY.

FT DISUFLID 304 325 BY SIMILARITY.

SQ SEQUENCE 354 AA; 40287 MW; 065D155378A1283C CRC64;

RESULT 9

PLK_CHICK

ID PLK_CHICK

STANDARD: PRT; 355 AA.

AC P07354;

DT 01-APR-1988 (Rel. 07, Created)

DT 01-APR-1988 (Rel. 07, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE PROTEOGLYCAN LINK PRECURSOR (CARTILAGE LINK PROTEIN) (LP).

GN CRTL1.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE="EMBRYONIC STERNAL CARTILAGE"; PubMed=459124;

RA MEDLINE=86223315; PubMed=459124;

RA Deak F., Kiss I., Sparks K.J., Argraves W.S., Hampikian G., Goetinck P.F.;

RT "Complete amino acid sequence of chicken cartilage link protein deduced from cDNA clones."

RL Proc. Natl. Acad. Sci. U.S.A. 83:3766-3770(1986).

RN [2]

RP SEQUENCE FROM N.A.

RC MEDLINE=87317659; PubMed=3476955;

RA Kiss I., Deak F., Mesrice S., Delius H., Soos J., Dekany K., Argraves W.S., Sparks K.J., Goetinck P.F.

RT "Structure of the chicken link protein gene: exons correlate with the protein domains."

RL Proc. Natl. Acad. Sci. U.S.A. 84:6399-6403 (1987).

CC -!- FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.

CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; M35038; AAA48941.1; JOINED.

DR EMBL; M35038; AAA48941.1; JOINED.

DR EMBL; M35038; AAA48941.1; JOINED.

DR EMBL; M13212; AAA48940.1; -.

DR PIR; A24881; LKCH.

DR PIR; A28305; A28305.

DR HSSP; P98066; 1TSG.

DR INTERPRO; IPR000538; -.

DR INTERPRO; IPR003006; -.

DR PFAM; PF00193; Xlink; 2.

DR PFAM; PF00047; ig; 1.

DR PROTEPE; PS01241; LINK; 2.

KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage; Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.

FT PROTEP 1 15 BY SIMILARITY.

FT CHAIN 16 354 BY SIMILARITY.

FT DOMAIN 54 146 IG-LIKE V-TYPE DOMAIN.

FT DOMAIN 176 253 LINK 1.

FT DOMAIN 274 350 LINK 2.

FT DISUFLID 61 139 BY SIMILARITY.

FT DISUFLID 181 252 BY SIMILARITY.

FT DISUFLID 205 226 BY SIMILARITY.

FT DISUFLID 279 349 BY SIMILARITY.

FT DISUFLID 304 325 BY SIMILARITY.

SQ SEQUENCE 354 AA; 40287 MW; 065D155378A1283C CRC64;

Query Match 8.9% Score 165.5; DB 1; Length 354;

Best Local Similarity 30.1% Pred. No. 1.9e-07;

Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;

FT CHAIN 10 355 PROTEOGLYCAN LINK PROTEIN.

FT DOMAIN 54 147 IG-LIKE V-TYPE DOMAIN.

FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	387	387	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	CARBOHYD	1675	1675	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT	SITE	1171	1173	CELL ATTACHMENT SITE (POTENTIAL).	CC
SQ	SEQUENCE	2132 AA;	222008 MW;	QB2BCDFC6CBDA163 CRC64;	CC
Query Match		8.9%	Score 165;	DB 1; Length 2132;	DR
Best Local Similarity		28.8%	Pred No. 1.9e-06;		EMBL: X17405; CAA35462.1; -;
Matches 49;	Conservative	23;	Mismatches 66;	Indels 32;	DR
				Gaps 5;	DR
Qy	57	VFHLRSPLGOKLTFDKAREACANEATATMAYNQSLSYKQAKYHLCAGWLETGRVAYPT	116		DR
Db	488	VFHYPGSIVRSLSIVFEAAQACMHTGATIASPEQLQAYAEYEQCDAGWLDQTYRPI	547		INTERPRO: IPR000538; -;
Qy	117	AFASRNC--GSGVGVGIVDYGPRPNKSEMWVDFCYRKMDVNTCTXKIVYVGDFSYSGNLL	173		INTERPRO: IPR003006; -;
Db	548	VSPRTPCIVGDKDSSPGVTRVPS-STTYDVCYVDK-----LEGIVF-----	590		PIR: S14944; LKHU;
Qy	174	QVLMSFPSLTINFELTEVLAYSNSSARGRAFHLTDLSITRGFLXPQNSGL	223		PFAM: PF00047; 1g; 1;
Db	591	-----FAFRLQEQTFQPARAFCAAQNAT-LASTGQLYAAWSQGL	628		PROSITE: PS01241; LINK 2.
RESULT	12				KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage; Immunoglobulin domain; repeat.
PLK_HUMAN					KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; repeat.
ID	PLK_HUMAN				FT PROPEP 15
AC	P10155;				FT PROPEP 1
DT	01-JUL-1989 (Rel. 11, Created)				FT CHAIN 16
DT	01-APR-1990 (Rel. 14, Last sequence update)				FT DOMAIN 54
DT	01-NOV-1997 (Rel. 35, Last annotation update)				FT DOMAIN 176
DE	PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP)				FT DOMAIN 274
GN	CRTL1.				FT DISULFID 61
OS	Homo sapiens (Human).				FT DISULFID 139
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				FT DISULFID 252
RN	[1]				FT DISULFID 226
RP	SEQUENCE FROM N.A.				FT DISULFID 279
RC	TISSUE=ARTICULAR CHONDROCYTES;				FT DISULFID 349
RX	Medline=90206798; PubMed=2320422;				FT DISULFID 325
RA	Dudhia J.; Hardingham T.E.;				FT CARBOHYD 21
RT	"The primary structure of human cartilage link protein.";				FT CARBOHYD 56
RL	Nucleic Acids Res. 18:1292-1292(1990).				FT CARBOHYD 354 AA; 315C96EC3AC2626A CRC64;
RN	[2]				Query Match 8.8%; Score 163.5; DB 1; Length 354;
RP	Medline=90245703; PubMed=2336413;				Best Local Similarity 30.1%; Pred. No. 2.8e-07;
RA	Dudhia J.; Hardingham T.E.;				Matches 44; Conservative 21; Mismatches 52; Indels 29; Gaps 4;
RL	Nucleic Acids Res. 18:2214-2214(1990).				Qy 5 GKHKCBKSHYVGGLNCEPEQLPIDRCLQDNQCHADAKCVDLHFQDTTVGVFLRSPL 64
RN	[3]				Db 135 GRYKE-----VTEGL-----EDDTVVVFLDQGV--VFPYFPRL 167
RP	SEQUENCE FROM N.A.				Qy 65 GQYKLTFLDKARECANEAATATMAYNOLSYXQKAKYHLCAGMLETGRAVATTAFAQNCG 124
RC	TISSUE=LARYNGEAL CARTILAGE CHONDROCYTES;				Db 168 GRYNLFHEHQQAQLDQDAVIAFSDLYDAWRGGLDWNAGWLSDGQVQYPTKPREPCG 227
RX	Medline=90206798; PubMed=2320422;				Qy 125 --SGVGVTVDYGPRPNKSEMWDYFCY 148
RA	Dudhia J.; Hardingham T.E.;				Db 228 QNNTPGVNRNGFNDKDSRYDVECF 253
RT	"Characterization of the promoter for the rat and human link protein gene.";				RESULT 13
RL	Nucleic Acids Res. 19:1933-1939(1991).				PLK_PTG STANDARD; PRT; 354 AA.
RX	Medline=91139126; PubMed=2286376;				ID PLK_PTG AC P10839; 11, Created)
RA	Osborne-Lawrence S.L., Sinclair A.K., Hicks R.C., Lacey S.W.,				DT 01-JUL-1989 (Rel. 11, Last sequence update)
RA	Eddy R.L. Jr., Byers M.G., Shows T.B., Duby A.D.;				DT 01-DEC-1998 (Rel. 37, Last annotation update)
RT	"Complete amino acid sequence of human cartilage link protein (CRTL1) deduced from cDNA clones and chromosomal assignment of the gene.";				DE PROTEOGLYCAN LINK PROTEIN PRECURSOR (CARTILAGE LINK PROTEIN) (LP).
RL	Genomics 8:562-567(1990).				GN CRTL1.
RN	[4]				OS sus scrofa (Pig).
RP	SEQUENCE FROM N.A.				OC Eukaryota; Chordata; Craniata; Vertebrata; Suina; Suidae; Sus.
RA	Rhodes C., Savigner P., Line S., Sasaki M., Chirigos M., Doege K.,				OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RA	Yamada Y.;				RN SEQUENCE FROM N.A.
RT	"Characterization of the promoter for the rat and human link protein gene.";				RC TISSUE=LARYNGEAL CARTILAGE CHONDROCYTES;
RT	Nucleic Acids Res. 19:1933-1939(1991).				RX Medline=8029337; PubMed=316916.
CC	-I - FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.				RA Peoples S.J., Neals A.S., Dudhia J., Hardingham T.E.;
CC	-I - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN FOLD AND TANDEM REPEAT STRUCTURES IN PROTEOGLYCAN N-				RT terminal domains and link protein.;"
CC	-I - SIMILARITY: CONTAINS 2 LINK DOMAINS.				RL J. Mol. Biol. 206:737-753(1989).

-1: FUNCTION: STABILIZE THE AGGREGATES OF PROTEOGLYCAN MONOMERS WITH
CC HYALURONIC ACID IN THE EXTRACELLULAR CARTILAGE MATRIX.
CC -1: SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1: SIMILARITY: CONTAINS 2 LINK DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch>).

CC EMBL; Y00165; CAA68358.1; -.
DR PIR; S04243; S04243.
DR HSSP; P98066; 1TSG.
DR INTERPRO; IPR005358; -.
DR INTERPRO; IPR030066; -.
DR PFAM; PF00193; Xlink; 2.
DR PFAM; PF00047; ig; 1.
PROTEIN; PS01241; LINK; 2.
KW Glycoprotein; Connective tissue; Extracellular matrix; Cartilage;
KW Hyaluronic acid; Proteoglycan; Immunoglobulin domain; Repeat.
FT PROTEIN 1 9
FT DOMAIN 10 354 PROTEOGLYCAN LINK PROTEIN.
FT DOMAIN 54 146 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 176 253 LINK 1.
FT DOMAIN 274 350 LINK 2.
FT DISULFID 61 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 349 BY SIMILARITY.
FT DISULFID 304 325 BY SIMILARITY.
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 354 AA; 40260 MW; 68FFB9DE51ABC01 CRC64;

Query Match 8.7%; Score 161.5; DB 1; Length 354;

Best Local Similarity 29.5%; Pred. No. 4.3e-07;
Matches 43; Conservative 22; Mismatches 52; Indels 29; Gaps 4;

QY 5 GKHCECKSHYYGDLNCEPEQQLPDRCLQNGQCHAACKYDHLFQDTIVGFVHLRSPL 64
Db 135 GRYCKE-.-.VIEGL-.-.EDDTAVVALLDEGV-.-VFVYFVPLR 167
QY 65 GQYKLTEDKARECANEAATMAYNOLSYQKAKYHLCASGWLLETGRVAYTAFASQNCG 124
Db 168 GRYNINFEHQAAQACLDQDIAVSFDQLYDAWRGGLDWCAGWLSDGSVQPTKPREPCG 227
QY 125 -SCYVGIVDYGPRPKNSEMWDFCY 148
Db 228 GONTVPGVNYGFWDKDRSDYDFVCF 253

RESULT 14
PGCA_HUMAN STANDARD, PRT; 2415 AA.
ID PGCA_HUMAN STANDARD,
AC P16112; Q13650;
DT 01-APR-1990 (Rel. 1.4, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE PROTEIN) (CSPCP) (CHONDRITIN SULFATE PROTEOGLYCAN CORE PROTEIN). AGC1 OR CSPG1.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1] SEQUENCE OF 1-2162; 2201-2415 FROM N.A.
RESEQUENCE OF 1-2162; 2201-2329 AND 2391-2415 FROM N.A.
RC TISSUE=CHONDROCYTES;
RN MEDLINE=91093289; PubMed=1965970;
RN Doege K.J.; Sasaki M.; Kimura T.; Yamada Y.;

"Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms.";
CC J. Biol. Chem. 266:894-902(1991).
RN [2] SEQUENCE OF 1778-2162 AND 2201-2415 FROM N.A.
RC TISSUE=CHONDROCYTES;
RA Duthia J.; Hardingham T.E.; Submitted (JAN-1990) to the EMBL/GenBank/DDJB databases.
RN [3] SEQUENCE OF 1936-2415 FROM N.A.
RP MEDLINE=89301054; PubMed=789216;
RX SEQUENCE OF 764-864 FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE=95128522; PubMed=1827755;
RT Barry F.-P.; Neame P.-J.; Sasse J.; Pearson D.;
RT "Length variation in the keratan sulfate domain of mammalian aggrecan";
RL J. Biol. Chem. 264:15747-15750(1989).
RN [4] TISSUE=BLOOD;
RX Matrix Biol. 14:323-328(1994).
-1: FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
CC -1: SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).
CC -1: ALTERNATIVE PRODUCTS: THERE ARE AT LEAST THREE FORMS OF AGGREGAN TRANSCRIPTS GENERATED BY ALTERNATIVE EXON USAGE. THE SHORTEST AND ALSO THE MOST PREVALENT FORM, LACKS BOTH THE EGF-LIKE DOMAIN, AND SUSHI DOMAIN.
CC -1: DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMprise THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDRITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
CC -1: PTM: CONTAINS MOSTLY CHONDRITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1: SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1: SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1: SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1: SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC -1: SIMILARITY: BELONGS TO THE AGGREGAN/VERSTICAN PROTEOGLYCAN FAMILY.

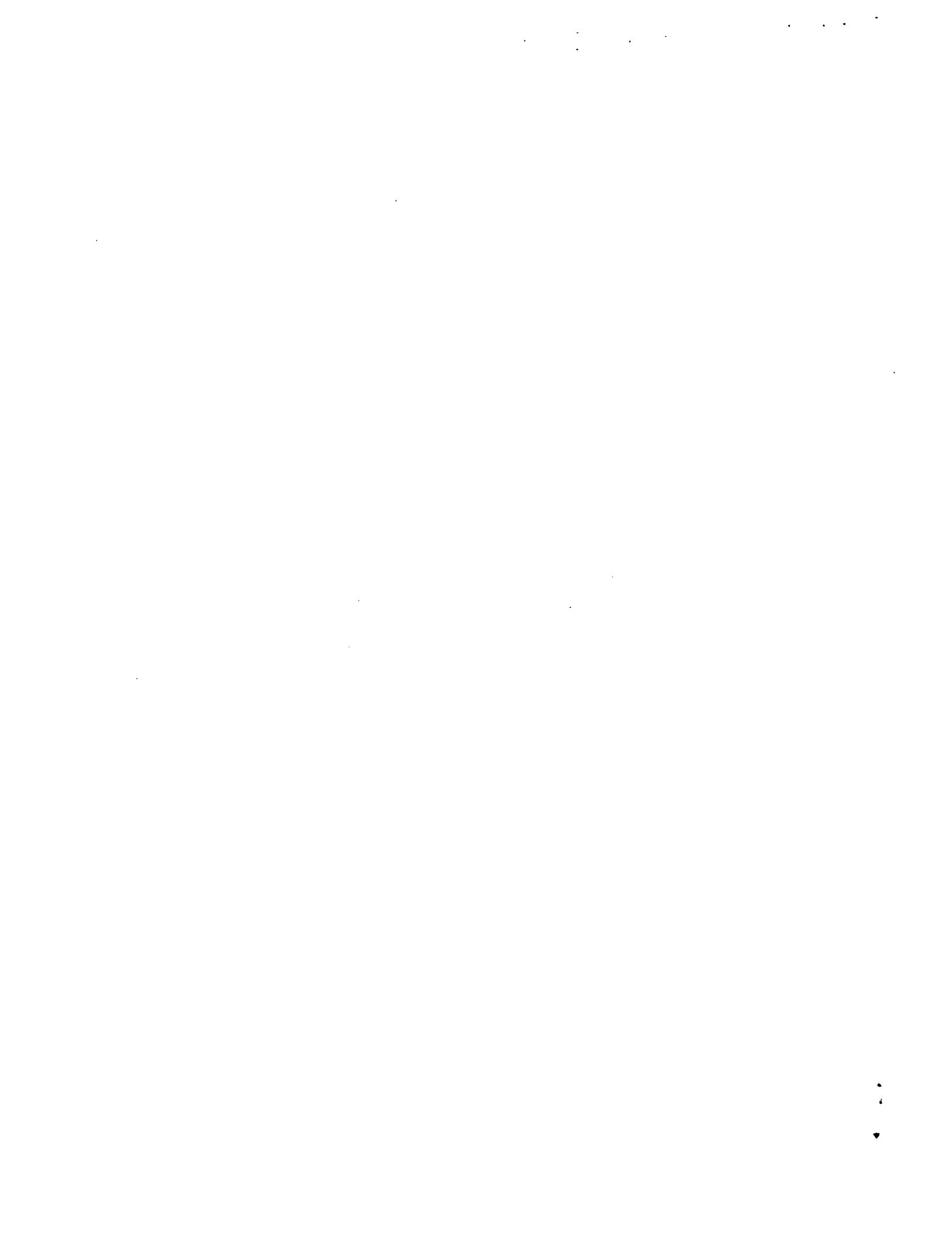
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch>).

CC DR EMBL; M5172; AAA2824.1; -.
CC DR EMBL; J05062; AAA5726.1; -.
CC DR EMBL; X1706; CAA3563.1; -.
CC DR EMBL; S74659; AAC0643.1; -.
CC DR PIR; S08042; S08042.
CC DR HSSP; P98066; 1TSG.
CC DR MIM; 155760; -.
CC DR INTERPRO; IPR000436; -.
CC DR INTERPRO; IPR000495; -.
CC DR INTERPRO; IPR000538; -.
CC DR INTERPRO; IPR000561; -.
CC DR PFAM; PF00008; EGF; 1.
CC DR PFAM; PF00193; Xlink; 4.

DR PROSITE; PS50041; C_TYPELECTIN_2; 1.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 EGF-like domain; Repeat; Lectin; Sushi; Signal.
 PT SIGNAL 1 22 POTENTIAL.
 PT CHAIN 23 1268 NEUROCAN CORE PROTEIN.
 PT DOMAIN 51 146 IG-LIKE V-TYPE DOMAIN.
 PT DOMAIN 176 253 LINK 1.
 PT DOMAIN 274 355 LINK 2.
 PT DOMAIN 960 996 EGF-LIKE 1.
 PT DOMAIN 998 1034 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 PT DOMAIN 1036 1165 C-TYPE LECTIN.
 PT REPEAT 1166 1224 SUSHI.
 PT DISULFID 181 139 BY SIMILARITY.
 PT DISULFID 205 252 BY SIMILARITY.
 PT DISULFID 205 226 BY SIMILARITY.
 PT DISULFID 279 354 BY SIMILARITY.
 PT DISULFID 303 324 BY SIMILARITY.
 PT DISULFID 964 975 BY SIMILARITY.
 PT DISULFID 969 984 BY SIMILARITY.
 PT DISULFID 986 995 BY SIMILARITY.
 PT DISULFID 1040 1051 BY SIMILARITY.
 PT DISULFID 1068 1160 BY SIMILARITY.
 PT DISULFID 1136 1152 BY SIMILARITY.
 PT DISULFID 1167 1210 BY SIMILARITY.
 PT DISULFID 1196 1223 BY SIMILARITY.
 PT CARBOYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
 PT CARBOYD 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1268 AA; 137200 MW; 3014E8E202A2FAEC CRC64;

Query Match 8.5%; Score 158.5; DB 1; Length 1268;
 Best Local Similarity 34.2%; Pred. No. 3.8e-06;
 Matches 41; Conservative 219; Mismatches 53; Indels 7; Gaps 3;
 Qy 47 DLHFQDTTVGVFHPLASPLGOVKKLTFDKARBCANEAATMAYNOLSYQAKVHLC5AGW 106
 Db 150 DLVTFLEVTVGHVYHAAARDRYALTFPEAQEACRLSSATIAAPRLQAAFFEDGFDNCDAGW 209
 Qy 107 LETGRVAYPTAFASONC--GSGVGVGIVDYGPRPNKSEMWDFCYRMKDYNCTXKVGYVG 163
 Db 210 LSDRTVRYPTIQSRPGCYGDRSSLPGVRSYG-RRDPQELDVICFARE--LGGEVVFYVG 265

Search completed: April 4, 2001, 13:06:41
 Job time: 140 sec



Gencore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:03:57 ; Search time 17.49 Seconds
(without alignments)
1370.436 Million cell updates/sec

Title: US-09-466-778-11
1 MTGPQKHKOCCKSHYVGDL.....ALAAAYSFRINKTIGFXH 353
Sequence:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_66:*

1: Pir1;*

2: Pir2;*

3: Pir3;*

4: Pir4;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1468	79.1	1069	2 T42681	hypothetical prote tumor necrosis fac
2	264	14.2	275	2 A47230	TSG-6 homolog PS4
3	262	14.1	277	2 A41735	hyaluronate-bindin
4	175	9.4	2109	1 I50421	agrecan precursor
6	171	9.2	2124	1 A28452	proteoglycan core
7	170	9.2	2327	2 T42630	agrecan - bovine
8	169.5	9.1	408	1 LKRT2	proteoglycan link
9	166.5	9.0	354	1 S42938	proteoglycan link
10	165.5	8.9	355	1 LKCH	proteoglycan link
11	165	8.9	2132	1 A55182	agrecan precursor
12	163.5	8.8	354	1 LKHU	proteoglycan link
13	161.5	8.7	354	1 S04243	proteoglycan link
14	161	8.7	2316	1 A39086	agrecan precursor
15	160	8.6	370	2 S29139	agrecan - pig (fr
16	158.5	8.5	1268	2 S52781	neurocan - mouse
17	157.5	8.5	371	2 A53908	brevican precursor
18	157.5	8.5	883	2 S49126	brevican precursor
19	157.5	8.5	1257	2 S28764	neurocan precursor
20	155.5	8.4	912	2 A54423	brevican precursor
21	154.5	8.3	862	2 S43932	versican - pig-tai
22	152.5	8.2	3562	2 A47171	chondroitin sulfate
23	151.5	8.2	1643	2 T14274	versican precursor
24	151.5	8.2	3381	2 T42369	versican precursor
25	150.5	8.1	883	2 S5753	brevican precursor
26	150.5	8.1	2409	1 A60979	versican precursor
27	148	8.0	2397	1 A55535	brevican precursor
28	147	7.9	378	2 I46268	versican precursor
29	139.5	7.5	1340	2 A39808	proteoglycan core

RESULT 2

JC6506

tumor necrosis factor stimulated gene-6 protein - mouse

30	122	6.6	113	2 A55885	chondroitin sulfate
31	111.5	6.0	362	2 A35616	T-cell surface gly
32	109.5	5.9	358	2 A34907	plasma membrane gly
33	109.5	5.9	363	2 A37009	CD44 homolog membr
34	108	5.8	395	2 T77371	CD44R5 - human
35	106	5.7	359	2 S24240	lymphocyte surface
36	105	5.7	1584	2 T22674	hypothetical prote
37	104	5.6	3507	2 T34513	hypothetical prote
38	103.5	5.6	365	2 A34424	CD44 membrane glyc
39	100	5.4	351	2 S43305	CD44 antigen precu
40	99.5	5.4	728	2 I50719	C-delta-1 - chick
41	99.5	5.4	3623	2 T0456	intrinsic factor-B
42	99	5.3	601	2 T2025	hypothetical prote
43	99	5.3	2907	2 A57278	fibrillin-2 precursor
44	98	5.3	361	2 G02251	lymphocyte surface
45	98	5.3			cell surface glyco

ALIGNMENTS

RESULT 1					
T42681	hypothetical protein DKFZp434E0321.1	-	human (fragments)		
C;Species: Homo sapiens (man)					
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000					
C;Accession: TA2681					
R;Blum, H.; Bauersachs, S.; Meves, H.W.; Gassendorfer, J.; Wiemann, S.					
A;Submitted to the Protein Sequence Database, November 1999					
A;Reference number: Z22233					
A;Accession: T42681					
A;Status: preliminary					
A;Molecule type: mRNA					
A;Residues: 1-870-871-1069 <AAA>					
A;Cross-references: EMBL:ALL3301					
A;Experimental source: adult testis; clone DKFZp434E0321					
A;Note: the cDNA sequence contains a +1 frameshift near codon 870					
C;Genetics:					
A;Note: DKFZp434E0321.1					
Query Match	Best Local Similarity	Score	DB 2;	Length	1069;
Matches 284; Conservative	80.5%	Pred. No. 1..1..-117;	4; Mismatches	29;	Indels 36; Gaps 2;
QY	1	MTPGKHCKCEKSHYVGDGLNEPEQIPIDCLQNGQCHADAKCVDFLQDFTVGVYH	60		
Db	697	MTPGKHCKCEKSHYVGDGLNEPEQIPIDCLQNGQCHADAKCVDFLQDFTVGVYH	756		
QY	61	RSPIGQYKLTDFKAREACANEATMATYNOLSYXKAKYHLCASAGWLEGRVAYPTAFAS	120		
Db	757	RSPIGQYKLTDFKAREACANEATMATYNOLSYXAKYHLCASAGWLEGRVAYPTAFAS	816		
QY	121	QNGSGVYVGIVDYGPDKNSKEMWDYFCYRMKDVNCTXKVYVGDGFSXSGNLQVLMSP	180		
Db	817	QNGSGVYVGIVDGPDKNSKEMWDYFCYRMKDVNCTXKVYVGDGFSXSGNLQVLMSP	860		
QY	181	SILNLTLEVILAYSNSSARGRAFLEHTLDSRGTLEPKQNSLGENETLSGRDLEHHLAN	240		
Db	861	CIS-----RTPDLSRTGTFVYQNSGLGENETLSGRDLEHHLAN	900		
QY	241	VSMEFFYNDLVNTTLPQRLGSKLLTDQDPLHPETRCYGRDTLWEDICASNGITHV	300		
Db	901	VSMEFFYNDLVNTTLPQRLGSKLLTDQDPLHPETRCYGRDTLWEDICASNGITHV	960		
QY	301	SRXIKAPAPVTLXHMJXGTFXXJLILVGTAVALAAYSFRINRKTIGXFH	353		
Db	961	SRPLKAPAPVTLTHTSGLGAGIFFAIIYGTAVALAAYSFRINRKTIGXFQHF	1013		
RESULT 2					
JC6506					
tumor necrosis factor stimulated gene-6 protein - mouse					

C;Species: *Mus musculus* (house mouse)
 C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C;Accession: JG6506
 R;Fueloep, C.; Kamath, R.V.; Li, Y.; Otto, J.M.; Salustri, A.; Olsen, B.R.; Glant, T.T.;
 Gene 202, 95-102, 1997
 A;Title: Coding sequence, exon-intron structure and chromosomal localization of murine T
 A;Reference number: JG6506; MUID:98087423
 A;Accession: JG6506
 A;Status: Preliminary
 A;Molecule type: mRNA
 A;Cross-references: GB:U83903; NID:92062474; PID: AAC53527.1; PID: 92062475
 Query Match
 Best Local Similarity 43.3%; Pred. No. 3.1e-15;
 Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;
 Qy 52 DTTGVYFHLSPLGQYKLTDFKAREACANEATMATYNQLSYXQKAYHLCAGWLETGR 111
 Db 32 EQAACVYHREARGVKLYTAAEKAVCEFEGRGATYKQLEARKIGFHVCAAGWAKGR 91
 Qy 112 VAYPTAFASQNCGSGSVVGIVTDYGRPRNPKNSEMDVYCFTRMKDYN 155
 Db 92 VGYPIVKPGNCGFGKGTIDYGLRNLNSERWDAYCYNPHAKEC 135
 RESULT 3
 A47290 TSG-6 homolog PS4 - rabbit
 N;Alternate names: Probable cell growth and differentiation protein, 12K; serum-inducible
 C;Species: *Oryctolagus cuniculus* (domestic rabbit)
 C;Date: 03-Mar-1994 #sequence_revision 07-Jul-1995 #text_change 08-Oct-1999
 C;Accession: A48055; A47290
 R;Feng, P.; Liu, G.
 J. Biol. Chem. 268, 21453, 1993
 A;Reference number: A48055; MUID:94012707
 A;Accession: A48055
 A;Molecule type: mRNA
 A;Residues: 1-276 <PEN>
 A;Cross-references: GB:W86381; NID:9387866; PID:AAA03342.1; PID: 9387867
 A;Note: this report replaces the sequence from reference A47290 and shares no similarity
 J. Biol. Chem. 268, 9387-9392, 1993
 A;Title: Identification of a novel serum and growth factor-inducible gene in vascular smooth muscle
 A;Reference number: A47290; MUID:93252803
 A;Accession: A47290
 A;Molecule type: mRNA
 A;Residues: 'MGSNSRSQLTTEVALDGTICRKNSFLVTRTNGLYAAVLSGLNRRPGVTVQRTBGNFSSRHESSCAREPQLC
 A;Experiments: source: vascular smooth muscle cells
 A;Note: sequence extracted from NCBI backbone (NCBIN:131589, NCBI:131590)
 A;Note: the sequence reported here was found to be artificial, a hybrid molecule of two
 C;Superfamily: Ctr/Cis repeat homology; link protein repeat homology <LINK>
 F:53-128/Domain: link protein repeat homology <C1R2>
 F:135-244/Domain: Ctr/Cis repeat homology <C1R2>

RESULT 5
 A47290 TSG-6 homolog PS4 - rabbit
 N;Alternate names: cartilage chondroitin sulfate proteoglycan core protein
 C;Species: *Gallus gallus* (chicken)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jan-2000
 C;Accession: 150421; S39796; S27356; A25442; A32002; 150216; A37072; B37002
 R;Li, H.; Schwartz, N.B.; Verteil, B.M.
 J. Biol. Chem. 268, 23504-23511, 1993
 A;Title: cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein
 A;Reference number: A48884; MUID:94043149
 A;Accession: 150421
 A;Status: preliminary; translated from GB/EMBL/DDJB
 A;Molecule type: mRNA
 A;Residues: 1-2109 <LINK>
 A;Cross-references: GB:L21913; NID:9416133; PID: AAB19128.1; PID: 9416134
 R;Chandrasekaran, L.; Tanzer, M.L.
 Biochem. J. 296, 895-887, 1993
 A;Reference number: S39796; MUID:94107258
 A;Contents: annotation; erratum
 A;Accession: S39796
 A;Molecule type: mRNA
 A;Residues: 1-361; DL, '364-600, 'R', 602-999, 'R', 1001-1028, 'P', 1030-1250, 'D', 1252-1602,
 A;Cross-references: GB:R88101
 R;Chandrasekaran, L.; Tanzer, M.L.
 Biochem. J. 288, 903-910, 1992
 A;Title: Molecular cloning of chicken aggrecan. Structural analyses.
 A;Reference number: S27356; MUID:93111968
 A;Accession: S27356
 A;Molecule type: mRNA

RESULT 4
 A41735 hyaluronate-binding protein TSG-6 precursor - human

A;Residues: 1-361; 'DL', '364-600, 'R', '602-999, 'R', '1001-1028, 'P', '1030-1250, 'D', '1252-1549, 'T'	Qy 187 TEVLAYSNSSARGRAFLEHLTDL-----SIRGTLFXPQNSGLGE-----
R;Sai, S.; Tanaka, T.; Kosher, R.A.; Tanzer, M.L.	Db 426 426 - -NETLSGRDIEHHHLANVS 242
Proc. Natl. Acad. Sci. U.S.A. 83, 5081-5085, 1986	Db 483 483 TPGLESASPFTEDHLVQVT 502
A;Title: Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Reference number: A25442; MUID:86259736	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Accession: A25442	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Molecule type: mRNA	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Cross-references: GB:MI:3993; NID:921654; PIDN:AAA48720.1; PID:9211655	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Experimental source: sternal cartilage	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
R;Tanaka, T.; Har-El, R.; Tanzer, M.L.	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
J. Biol. Chem. 263, 15831-15835, 1988	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Title: Partial structure of the gene for chicken cartilage proteoglycan core protein.	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Reference number: A32002; MUID:89008500	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Accession: A32002	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Molecule type: DNA	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Cross-references: 1893-1987-'S', 1989-2022 <TAN>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Note: the authors translated the codon TCC for residue 1787 as Phe	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
R;Krueger, R.C.	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
J. Biol. Chem. 265, 12098-12097, 1990	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Title: Chick cartilage chondroitin sulfate proteoglycan core protein: II. Nucleotide sequence	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Reference number: 150216; MUID:90307744	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Accession: 150216	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Molecule type: mRNA	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Residues: 'PA', 1044-1559 <KR2>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Cross-references: GB:MI:38187; NID:9211685; PIDN:AAA8731.1; PID:9555441	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
R;Krueger Jr., R.C.; Fields, T.A.; Hildreth IV, J.; Schwartz, N.B.	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
J. Biol. Chem. 265, 12098-12087, 1990	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Title: Chick cartilage chondroitin sulfate proteoglycan core protein. I. Generation and characterization	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Reference number: A37072; MUID:90307743	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Accession: A37072	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Molecule type: protein	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Residues: 998-1015, 'X', 1017-1019, 'X', 1021-1023 <KR2>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Note: amino end of 86K core Peptide CS-A	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Accession: B37072	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Molecule type: protein	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Residues: 1247-1250, 'D', 1252-1272, 'X', 1274-1275 <KR3>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Note: amino end of 75K core Peptide CS-B	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
A;Superfamily: Iggreen; C-type lectin homology; complement factor H repeat homology; EC	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
C;Keywords: alternative splicing	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;21-120/Domain: signal sequence #status predicted <SIG>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;21-2109/Domain: aggreen #status predicted <MR1>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;44-131/Domain: immunoglobulin homology <IMM>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;166-243/Domain: link protein repeat homology <LNK1>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;264-346/Domain: link protein repeat homology <LNK2>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;537-614/Domain: link protein repeat homology <LNK3>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;1856-1890/Domain: EGF homology <EGF>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;1891-2017/Domain: C-type lectin homology <LCH>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
F;2024-2080/Domain: complement factor H repeat homology <PHD>	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.4%; Score 175; DB 1; Length 2109;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 23.4%; Pred. No. 1.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 75; Conservative 31; Mismatches 116; Indels 98; Gaps 12;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 10 ECKSHYVGDLNCEPEQLPDIRCLQDNGOCHA-----DAKCVDLHFDQDTVG 56	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 194 QCDAGWLAQDQTVPIHLPRERCYIHKDKEFQYRTGYRETDYTYCYAEQMQGK--- 250	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VFLHRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 251 VFYATSP---EKEFIFQEAQDCKHSIGLARLATGELYLAWKDMCAGMDSAGWLADRVSRYPI 307	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASONCGSGVGVI---VDYGRP---PKNSEMWDVFCYRMKDVNCPTKVGTVGDSYSGNL 162	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 308 SRAPNCGGNIVGVRVYLNPAQNTGYPHPSRSRDAICVSDDFEALVPLGFTDEVGTEL 367	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 163 GDGFSY-----SSGNLILQYLMSPFSLTNFL 186	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 368 GSAFTIQTQVQEYELPLPRLPRNVTBEEARGSIATATEPMEITATALEYBAFTVLPDL--FA 425	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 548 VSPRTCPVGDKDSSPGVRYTGVPRS-SETDVYCVYDVK-----LEGEVF----- 590	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Query Match 9.2%; Score 171; DB 2; Length 2124;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Best Local Similarity 28.8%; Pred. No. 3.7e-06;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Matches 49; Conservative 24; Mismatches 65; Indels 32; Gaps 5;	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 57 VPHLRSPLGQYKLFEDKARECANAAATMAYNOLSYQAKYHLCASQWLITGRVAYPT 116	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Db 488 VPHYRPGSTRSLTFEAQQACIRGAAIASPEQLQAAEAGYVQCDAGWLQDQTVPI 547	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628
Qy 117 AFASQNSQ---GSGVYVIVDVGPRNPKSEMWDFVCFYRMKDVNCPTKVGTVGDSYSGNL 173	Db 591 591 - - - - - FATQMEQFTFQEAQFCAAQNAT - LASTGQQLYAAWSQQL 628

RESULT 7

T42630 - bovine
agrecan - Bos primigenius taurus (cattle)
C;Species: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C;Accession: T42630
R;Hesing, T.M.; Kollar, J.; Hunyh, T.D.
A;Description: Complete coding sequence of bovine aggrecan: comparative structural analysis
A;Accession: T42630
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Residues: 1-2327 <HERB>
A;Cross-references: EMBL:U76615; NID:91730239; PID:91730260; PIDN:AAB38544.1
A;Experimental source: articular chondrocytes
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; Ig glycoprotein
C;Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein

Query Match 9.2%; Score 170; DB 2; Length 2327;
Best Local Similarity 38.9%; Pred. No. 5e-06;
Matches 37; Conservative 12; Mismatches 42; Indels 4; Gaps 2;

Qy 57 VFHLRSPLGQKLTDKARECANAAATMATYNOLSYQKAKYHLCSSAGWLETGRVAYPT 116
Db 488 VFHYRPGSSSYSLIFEEKACRLTGAIISPEQLQAAVEAGYEQCDAGWLDQDVTYPI 547

Qy 117 AFASQNC--GSGIVGIVDVGPRPKNSEMWDFVCFY 148
Db 548 VSPRTPCVGDKDSSPFGVRP-PSEFTDVVCFY 581

RESULT 9

S42938 9
proteoglycan link protein precursor - horse
N;Alternate names: cartilage link protein
C;Species: Equus caballus (domestic horse)
C;Accession: S42938
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 22-Jun-1999
R;Dudhia, J.; Platt, D.
submitted to the EMBL Data Library, March 1994
A;Reference number: S42938
A;Accession: S42938
A;Molecule type: mRNA
A;Residues: 1-354 <DDU>
A;Cross-references: EMBL:X78077; NID:9459438; PID:9459439
C;Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C;Keywords: cartilage; duplication; extracellular matrix; glycoprotein
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-354/Product: proteoglycan link protein #status predicted <MAT>
F;154-141/Domain: immunoglobulin homology <IMM>
F;116-25/Domain: link protein repeat homology <LNK1>
F;274-350/Domain: link protein repeat homology <LNK2>
F;21,56/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;61-139/disulfide bonds: #status predicted

Query Match 9.0%; Score 166.5; DB 1; Length 354;
Best Local Similarity 30.1%; Pred. No. 9.7e-07;
Matches 44; Conservative 22; Mismatches 51; Indels 29; Gaps 4;

Qy 5 GKKCCEKSHYVGDLNCEPEQLDRCLDNGQOCHADAKYCDLHFQDTTVGVFHLRSPL 64
Db 135 GRYKCE---VIEGL-----EDDTAVALDLQGV--VFYFPR 221

Qy 65 GQYKLTFTDKARECANAAATMATYNOLSYQKAKYHLCSSAGWLETGRVAYPTAFSQNC 124
Db 222 GRYLNHHEAQACLDDDAVASEFDQLYDAARGGLDWCNCAGWLSDGSVQSPITKPREPCG 281

Qy 125 --SGVVGIVDVGPRPKNSEMWDFVCFY 148
Db 282 GQNTVPGVRYNGFWDKDKSRTDVFCF 307

RESULT 10

LKRT2
proteoglycan link protein 2 precursor - rat
N;Alternate names: cartilage link protein
C;Species: Rattus norvegicus (Norway rat)
C;Accession: 13-Aug-1986 #sequence_revision 24-Oct-1997 #text_change 13-Nov-1998
R;Rhodes, C.; Doege, K.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 263, 6063-6067, 1988
A;Title: Alternative splicing generates two different mRNA species for rat link protein.
A;Reference number: A28654; MUID:88198139
A;Accession: A28654
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-408 <RHO>
R;Doege, K.; Hassell, J.R.; Caterson, B.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 83, 3761-3765, 1986
A;Title: Link protein cDNA sequence reveals a tandemly repeated protein structure.
A;Reference number: A24880; MUID:86233314
A;Accession: A24880
A;Molecule type: mRNA
A;Residues: 180-408 <DOE>
R;Neame, P.J.; Christopher, J.E.; Baker, J.R.
J. Biol. Chem. 261, 3519-3535, 1986
A;Title: The primary structure of link protein from rat chondrosarcoma proteoglycan agar
A;Reference number: A02869; MUID:86140139
A;Accession: A02869
A;Molecule type: protein
A;Residues: 16-33, A' 89-375, 'W' 377-408 <NEA>
C;Comment: This protein was extracted from rat chondrosarcoma.
C;Comment: Residues 259-277 and 358-376 (approximately) form epitopes for the species-specific antibodies.
C;Comment: Link proteins interact with and stabilize aggregates of hyaluronic acid and collagen.
C;Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
C;Keywords: cartilage; chondroitin sulfate proteoglycan; duplication; glycoprotein; hyal
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-408/Product: proteoglycan link protein 2 #status experimental <MAT>
F;198-495/Domain: immunoglobulin homology <IMM>
F;230-307/Domain: link protein repeat homology <LNK1>
F;328-404/Domain: link protein repeat homology <LNK2>

RESULT 10

LKCH
proteoglycan link protein precursor - chicken
C;Species: Gallus gallus (chicken)
C;Accession: A2805; A24881; B39097
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987
R;Kiss, I.; Deak, F.; Mestric, S.; Delius, H.; Soos, J.; Dekany, K.; Argraves, W.S.;
Proc. Natl. Acad. Sci. U.S.A. 84, 6399-6403, 1987
A;Title: Structure of the chicken link protein gene: exons correlate with the protein
A;Reference number: A28305; MUID:87317659

A; Accession: A28305	A; Molecule type: DNA	A; Cross-references: GB:LO7049; NID:9678541; PID:9191772
A; Residues: 1-355 <KIS>		R.Watanabe, H.; Gao, L.; Yamada, Y.
A; Cross-references: GB: M35038; NID: 9212264; PID: AAA48941.1; PID: 9212267		Biochem. J. 308, 433-440, 1995
PROC. NATL. ACAD. SCI. U.S.A. 83, 3765-3770, 1986		Alt title: Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene structure, and expression in mouse skeletal tissues
A; Title: Complete amino acid sequence of chicken cartilage link protein deduced from cDNAs		A; Reference number: SS5329; MUID: 95289972
A; Accession: A24881; MUID: 86233315		A; Status: preliminary
A; Molecule type: mRNA		A; Molecule type: mRNA
A; Residues: 1-355 <DEA>		A; Residues: 1-567, 'G', 569-1728, 'I', 1730-2132 <WAT1>
A; Cross-references: GB: U222801; NID: 9886014		A; Cross-references: GB: U222801; NID: 9886014
R; Glumoff, V.; Savontos, M.; Vehorio, E.		R; Glumoff, V.; Savontos, M.; Vehorio, E.
A; Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues		Biochim. Biophys. Acta 1219, 613-622, 1994
A; Reference number: SS50206; MUID: 95035091		A; Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues
A; Accession: SS50207		A; Reference number: SS50206; MUID: 95035091
A; Molecule type: mRNA		A; Molecule type: mRNA
A; Residues: 350-481, 'R', 483-506 <GLU1>		A; Residues: 350-481, 'R', 483-506 <GLU1>
R; Glumoff, V.		R; Glumoff, V.
A; Title: Complete sequence of the EMBL Data Library, July 1994		submitted to the EMBL Data Library, July 1994
A; Reference number: SS11355		A; Reference number: SS11355
A; Status: preliminary		A; Status: preliminary
A; Molecule type: mRNA		A; Molecule type: mRNA
A; Residues: 350-363, 'CRMSSQERPVA', <GLU2>		A; Residues: 350-363, 'CRMSSQERPVA', <GLU2>
A; Cross-references: EMBL: X00279		A; Cross-references: EMBL: X00279
R; Watanabe, H.; Kimata, K.; Line, S.; Strong, D.; Gao, L.Y.; Kozak, C.A.; Yamada, Y.		R; Watanabe, H.; Kimata, K.; Line, S.; Strong, D.; Gao, L.Y.; Kozak, C.A.; Yamada, Y.
A; Title: Cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the agg		A; Title: Cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the agg
A; Reference number: 156123; MUID: 95004579		A; Reference number: 156123; MUID: 95004579
A; Status: preliminary		A; Status: preliminary
A; Molecule type: DNA		A; Molecule type: DNA
A; Accession: 178532		A; Accession: 178532
A; Status: preliminary; translated from GB/EMBL/DBJ		A; Status: preliminary; translated from GB/EMBL/DBJ
A; Cross-references: GB: S73722; NID: 9765215; PID: 9765216		A; Cross-references: GB: S73722; NID: 9765215; PID: 9765216
A; Molecule type: DNA		A; Molecule type: DNA
A; Residues: 211-326 <WAT2>		A; Residues: 211-326 <WAT2>
A; Cross-references: GB: S73722; NID: 9765215; PID: 9765216		A; Cross-references: GB: S73722; NID: 9765215; PID: 9765216
A; Accession: 158123		A; Accession: 158123
A; Status: preliminary; translated from GB/EMBL/DBJ		A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA		A; Molecule type: DNA
A; Residues: 211-240, 'MCTASLRRWRYVSFMRHPOQRNSPSRQPTS', 'AGGWGHAWPPQASSTWPGRAWTTCAGAGW'		A; Residues: 211-240, 'MCTASLRRWRYVSFMRHPOQRNSPSRQPTS', 'AGGWGHAWPPQASSTWPGRAWTTCAGAGW'
A; Cross-references: GB: S73720; NID: 9765211; PID: 9765211		A; Cross-references: GB: S73720; NID: 9765211; PID: 9765211
C; Genetics:		C; Genetics:
A; Introns: 34/1; 159/1; 260/1		A; Map position: 7
A; Note: single copy gene		A; Introns: 233/1
C; Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat		C; Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology
C; Keywords: cartilage; duplication; extracellular matrix; glycoprotein		C; Keywords: cartilage; extracellular matrix
F; 1-15 / Domain: signal sequence #status predicted <SIG>		C; Key words: cartilage; extracellular matrix
F; 16-35 / Domain: link protein #status predicted <MAT>		F; 1-19 / Domain: signal sequence #status predicted <SIG>
F; 54-142 / Domain: immunoglobulin homology <IMM>		F; 4-135 / Domain: immunoglobulin homology <IMM>
F; 177-254 / Domain: link protein repeat homology <LINK1>		F; 110-249 / Domain: link protein repeat homology <LINK1>
F; 275-351 / Domain: link protein repeat homology <LINK2>		F; 258-349 / Domain: link protein repeat homology <LINK2>
F; 21/21 Binding site: carbohydrate (Asn) (covalent) #status predicted		F; 54-581 / Domain: link protein repeat homology <LINK3>
F; 56/56 Binding site: carbohydrate (Asn) (covalent) #status experimental		F; 602-683 / Domain: link protein repeat homology <LINK4>
F; 61-140,182-253,206-227,280-350,305-326/Disulfide bonds: #status predicted		F; 1122-2042 / Domain: C-type lectin homology <LCH>
F; 2049-2105 / Domain: complement factor H repeat homology <FHD>		F; 2049-2105 / Domain: complement factor H repeat homology <FHD>
Query Match 8.9%; Score 165.5; DB 1; Length 355;		Query Match 8.9%; Score 165; DB 1; Length 2132;
Best Local Similarity 29.9%; Pred. No. 1.2e-06;		Best Local Similarity 28.8%; Pred. No. 1.2e-05;
Matches 44; Conservative 26; Mismatches 46; Indels 31; Gaps 5;		Matches 49; Conservative 23; Mismatches 66; Indels 32; Gaps 5;
Qy 5 GKHCKECKSHYVGDGQLNCEPEQLP1DRCLQDNGQCHADAKCYDHLFQDTIVGVFHLRSP- 63		Qy 57 VFHLRSPLQGQYKLTDFKARECANEAATMNTQNSYXQKAYHLCAGNLETGRVAYPTAFASNC 123
Db 136 GRYKE----VIGEL----EDTAVAVINLE----GTVFPKPR 167		Db 488 VFHVRPGSTRYSLTFFEEAQACAHITGAIASPEQQLQAYEYEQCDAGWLQDQTRYPI 547
Qy 64 LGQYKLTDFKARECANEAATMNTQNSYXQKAYHLCAGNLETGRVAYPTAFASNC 123		Qy 117 AFASONC --GSGVVGIVDYGPRPKNSMMDVYFCYRMKDVNCTXKVYVGDFSFYSGNLL 173
Db 168 LGRYNINFHAAQQACLQDQSTLIASDQYEAWSRGDLSQVQYPTKPRPC 227		Db 548 VSPRIPCVGDKDSSPGVRYVRPS-SETDYCYVKD-----LEGEVF-----590
Qy 124 G--SGWGVGIVDYGPRPKNSMMDVFCY 148		Qy 174 QVLMSEPSUTNTELTEVAYNSSSAAGRFALEHTDLSIRGTLFXPQNGL 223
Db 228 GGKNTVPGVNYGFMDKERSRYDFCCF 254		Db 591 -----FATRLEQTFQEARFAQAAQNT-LASTGOLIYAWSQGL 628
RESULT 11		
A55182		
aggrecan precursor - mouse		
C; Species: Mus musculus (house mouse)		
C; Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 21-Jan-2000		
C; Accession: A55182; S55329; S50201; S51355; T58123		
R; Walcz, E.; Deak, F.; Dehardt, P.; Coulter, S.N.; Fueloep, C.; Horvath, P.; Doege, K.J.		
Genomics 22, 364-371, 1994		
A; Title: Complete coding sequence, deduced primary structure, chromosomal localization,		
A; Reference number: A55182; MUID: 95104847		
A; Status: preliminary; nucleic acid sequence not shown; translation not shown		
A; Molecule type: mRNA		
A; Residues: 1-232 <WAT1>		

LKHU
 proteoglycan link protein precursor - human
 N;Alternate name: cartilage link protein
 C;Species: Homo sapiens (mna)
 C;Date: 31-Dec-1991 #text_change 22-Jun-1999
 C;Accession: S14914; S04244; S03868; A36308; S14926; S03509; S08041
 R;Budhia, J.; Hardingham, T. E.
 Nucleic Acids Res. 18, 1292-1299, 1990
 A;Title: The primary structure of human cartilage link protein.
 A;Reference number: S14914; MUID:90206798
 A;Accession: S14914
 A;Molecule type: mRNA
 A;Residues: 1-354 <DUD2>
 A;Residues: 223-354 <DUD2>
 A;Cross references: EMBL:Y00166
 A;Note: The authors translated the codon CTG for residue 264 as Arg and TTG for residue 265 as Cys. The authors note that the codon GAT for residue 93 is Ala.
 R;R;Nguyen, Q.; Murthy, G.; Roughtley, P.J.; Mort, J.S.
 J. Mol. Biol. 206, 749-753, 1989
 A;Title: Isolation and sequence of cDNA clones for pig and human cartilage link protein.
 A;Reference number: S04243
 A;Accession: S04243
 A;Molecule type: mRNA
 A;Residues: 1-354 <NGU>
 R;R;Osborne-Lawrence, S. L.; Sinclair, A.K.; Hicks, R.C.; Lacey, S.W.; Eddy Jr., R.L.; Byer
 Genomics 8, 56-567, 1990
 A;Title: Complete amino acid sequence of human cartilage link protein (CRTL1) deduced from the cDNA sequence. Evidence for proteoglycan aggregate by a cartilage metalloproteinase. Evidence
 A;Reference number: S03868; MUID:89246328
 A;Accession: S03868
 A;Molecule type: Protein
 A;Residues: 16-35 <NGU>
 R;Budhia, J.; Hardingham, T.E.
 Nucleic Acids Res. 18, 2212, 1990
 A;Title: The primary structure of human cartilage link protein.
 A;Reference number: S14926; MUID:90245703
 A;Accession: S14926
 A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
 A;Residues: 1-354 <OSB>
 R;Budhia, J.; Hardingham, T.E.
 Nucleic Acids Res. 18, 2212, 1990
 A;Title: The primary structure of human cartilage link protein.
 A;Reference number: S03868; MUID:89246328
 A;Accession: S03868
 A;Molecule type: mRNA
 A;Residues: 1-354 <NGU>
 A;Molecule type: mRNA
 A;Residues: 1-92, A', 94-354 <NGU>
 R;R;Nguyen, Q.; Mort, J.S.; Roughtley, P.J.
 J. Biol. Chem. 266, 569-573, 1990
 A;Title: Cartilage link protein is degraded more extensively by cathepsin L than by cathepsin B. The primary structure of the human cartilage link protein is deduced from the cDNA sequence. Evidence for proteoglycan aggregate by a cartilage metalloproteinase. Evidence
 A;Reference number: S09309; MUID:90197639
 A;Accession: S09309
 A;Molecule type: Protein
 A;Residues: 16-38, 'X', 40-55, 57-60, 62-65, 'X', 67, 'X', 69-80 <NG2>
 C;Genetic-c
 C;Genetic-c
 A;Gene: GDB:CRTL1
 A;Cross references: GDB:125232; OMIM:115435
 A;Map position: 5q13-5q14
 C;Superfamily: proteoglycan link protein; immunoglobulin homology; link protein repeat
 C;Keywords: cartilage duplication; extracellular matrix; glycoprotein
 C;Key words: 1-15/Domain: signal sequence #status predicted <S>
 F;F:1-15/Domain: proteoglycan link protein #status experimental <S>
 F;F:16-35/Domain: proteoglycan link protein #status predicted <S>
 F;F:17-62/Domain: immunoglobulin homology <IMM>
 F;F:27-35/Domain: link protein repeat homology <LNK1>
 F;F:56/Domain: link protein repeat homology <LNK2>
 F;F:61-139/Disulfide bonds: #status predicted

Sequence Comparison Results						
Query	Match	Score	Length	DB	Start	End
Qy	5 GKHKECKSHYVGDBLNCBPEQLPDIRCLQDNQCHADAKCVDLHFQDFTTVGFHRLRSP	8.7%	161.5	1;	Length	354
Db	135 GRYKEE----VIGEL-----	29.5%	260-06;	43;	Best Local Similarity	29.5%
Qy	65 GQYKLTDFKAREACANEAAATMAYNOLSYKAXYKSHCAGMLETGRVAYPTAFASQNCG	124	1;	1;	Matches	43;
Db	168 GRYNUNFHQAQOACDQDAVIAQSDFQDLYDAWGRGLDNCAGWLSGSQYPTTKPBP	148	1;	1;	Conservative	22;
Qy	125 -SGVWGTIVDYGPRPKNSEMWIDFCY	148	1;	1;	Mismatches	52;
Db	228 GQNTYPGVRYNYGFNDKDSRYDVFCC	253	1;	1;	Indels	29;
RESULT						
A39086	aggregate precursor	-	human			
N	Alternative names:	chondroitin sulfate proteoglycan 1;	large aggregating proteoglycan			
C	Species:	Homo sapiens (man)				
C	Date:	10-Sep-1999	#sequence_revision	10-Sep-1999	#text_change	28-Jan-2000
C	Accession:	A39086; S50206; A43919; S46659; S56389; S66646; B43919; C43919				

J. Biol. Chem. 266, 894-902, 1991
 A;Title: Complete coding sequence and deduced primary structure of the human cartilage 1
 A;Reference number: A39086; MUID: 91093289
 A;Accession: A39086
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-2316 <DDE>
 A;Cross-references: GB:M55172; NID:9178258; PID:AA62824.1; PID:9178259
 R;Jumhoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
 Biochim. Biophys. Acta 1219, 613-622, 1994
 A;Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues by
 A;Reference number: S50206; MUID: 95035091
 A;Accession: S50206
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 350-497 <GLU>
 A;Cross-references: ENBL:X80278; NID:9516295
 A;Note: this translation is not annotated in GenBank entry HSAGGREC, release 113.0
 A;Reference number: A43919; MUID: 92235266
 A;Accession: A43919
 A;Molecule type: protein
 A;Residues: 361-370, 'X', 372-373; 393-399, 'X', 401-407, 'X', 409 <SAN>
 A;Experimental source: synovial fluid
 A;Note: sequence modified after extraction from NCBI backbone
 Matrix Biol. 14, 323-328, 1994
 A;Title: Length variation in the keratan sulfate domain of mammalian aggrecan.
 A;Reference number: 146998; MUID: 95128522
 A;Molecule type: DNA
 A;Residues: 764-765, 'A', 767-846, 'V', 848-862, 'X', 864 <BAR>
 A;Cross-references: ENBL:S74659; NID:9807127; PID:AA0643.1; PID:9807128
 A;Note: the authors translated the codon GAA for residue 803 as Ala and CYT for residue
 R;Ilic, M.Z.; Mok, M.T.; Williamson, O.D.; Campbell, M.A.; Hughes, C.E.; Handley, C.J.
 Arch. Biochem. Biophys. 322, 22-30, 1995
 A;Title: Catsbtolism of aggrecan by explant cultures of human articular cartilage in the
 A;Reference number: S66389; MUID: 96004775
 A;Accession: S66389
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 17-23; 24, 'X', 26-27; 393-401; 402-403 <IL1>
 R;Fosang, A.J.; Last, K.; Knauper, V.; Murphy, G.; Neame, P.J.
 FEBS Lett. 380, 17-20, 1996
 A;Title: Degradation of cartilage aggrecan by collagenase-3 (MMP-13).
 A;Reference number: S66389; MUID: 96181659
 A;Accession: S66389
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 'V', 404-405, 'XX', <FO5>
 C;Genetics:

J. Biol. Chem. 266, 894-902, 1991
 A;Title: Complete coding sequence and deduced primary structure of the human cartilage 1
 A;Reference number: A39086; MUID: 91093289
 A;Accession: A39086
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-2316 <DDE>
 A;Cross-references: GB:M55172; NID:9178258; PID:AA62824.1; PID:9178259
 R;Jumhoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
 Biochim. Biophys. Acta 1219, 613-622, 1994
 A;Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues by
 A;Reference number: S50206; MUID: 95035091
 A;Accession: S50206
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 350-497 <GLU>
 A;Cross-references: ENBL:X80278; NID:9516295
 A;Note: this translation is not annotated in GenBank entry HSAGGREC, release 113.0
 A;Reference number: A43919; MUID: 92235266
 A;Accession: A43919
 A;Molecule type: protein
 A;Residues: 361-370, 'X', 372-373; 393-399, 'X', 401-407, 'X', 409 <SAN>
 A;Experimental source: synovial fluid
 A;Note: sequence modified after extraction from NCBI backbone
 Matrix Biol. 14, 323-328, 1994
 A;Title: Length variation in the keratan sulfate domain of mammalian aggrecan.
 A;Reference number: 146998; MUID: 95128522
 A;Molecule type: DNA
 A;Residues: 764-765, 'A', 767-846, 'V', 848-862, 'X', 864 <BAR>
 A;Cross-references: ENBL:S74659; NID:9807127; PID:AA0643.1; PID:9807128
 A;Note: the authors translated the codon GAA for residue 803 as Ala and CYT for residue
 R;Ilic, M.Z.; Mok, M.T.; Williamson, O.D.; Campbell, M.A.; Hughes, C.E.; Handley, C.J.
 Arch. Biochem. Biophys. 322, 22-30, 1995
 A;Title: Catsbtolism of aggrecan by explant cultures of human articular cartilage in the
 A;Reference number: S66389; MUID: 96004775
 A;Accession: S66389
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 17-23; 24, 'X', 26-27; 393-401; 402-403 <IL1>
 R;Fosang, A.J.; Last, K.; Knauper, V.; Murphy, G.; Neame, P.J.
 FEBS Lett. 380, 17-20, 1996
 A;Title: Degradation of cartilage aggrecan by collagenase-3 (MMP-13).
 A;Reference number: S66389; MUID: 96181659
 A;Accession: S66389
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 'V', 404-405, 'XX', <FO5>
 C;Genetics:

Query Match 8.7%; Score 161.5; DB 1; Length 2316;
 Best Local Similarity 27.9%; Pred. No. 2.7e-05;
 Matches 50; Conservative 22; Mismatches 64; Indels 43; Gaps 6;

Qy 47 DLHFQDITVG -----VFLHRSPLGQYKLTEDKARACANEATMATYNYNOLSYKOKA 97
 Db 460 DLVQVTAvgOPQHPLGGVYHFRGPTVSLTEEAQACPGTGVATASPEQLOQAYEA 519
 Qy 98 KYHLC3SAGWLETGRVAYPTAFASQNC--GSGVYVGIVDYGPRPKNSEMWDFVCY---- 148
 Db 520 GYEQCDAGWLBDQTVRPIVSPRTICVGKDSS-QVRLVGVRS-TEYDVCFVDLEG 578
 Qy 149 -----RMKD-----NETXKYGYYGDFGS-----YSGNLLQVLMSEPSLT 183
 Db 579 EYFFATRLEQFTFQEAELEFCBSHNTATTCQLYAAWSRGIDKCYAGNLDGSLRYPITV 637

RESULT 15
 S29139

aggrecan - pig (fragment)
 C;Species: Sus scrofa domestica (domestic pig)
 C;Date: 30-Sep-1993 #sequence revision 18-Mar-1997 #text_change 13-Aug-1999
 C;Accession: S29139; S32530; A44108
 R;Barry, F.P.; Gaw, J.U.; Young, C.N.; Neame, P.J.
 Biochem. J. 286, 761-769, 1992
 A;Title: Hyaluronan-binding region of aggrecan from pig laryngeal cartilage. Amino ac
 A;Reference number: S29139; MUID: 93038505
 A;Molecule type: protein
 A;Residues: 1-28, 'W', 30-370 <BAR1>
 R;Barry, F.P.; Gaw, J.U.; Young, C.N.; Neame, P.J.
 Biochem. J. 291, 951, 1993
 A;Title: Correction: Hyaluronan-binding region of aggrecan from pig laryngeal cartila
 A;Accession: S29139; MUID: 93256313
 A;Molecule type: protein
 A;Residues: 25-55 <BAR2>
 A;Molecule type: protein
 A;Accession: S32330
 A;Note: correction to sequence S29139
 R;Rosang, A.J.; Neame, P.J.; Last, K.; Hardingham, T.E.; Murphy, G.; Hamilton, J.A.
 J. Biol. Chem. 267, 19470-19474, 1992
 A;Title: The interglobular domain of cartilage aggrecan is cleaved by PUMP, gelatinas
 A;Reference number: A44108; MUID: 92406898
 A;Accession: A44108
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 345-348, 'NP', 349-351 <FO5>
 C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
 C;Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycop
 F;28-119/Domain: immunoglobulin homology <IMM>
 F;154-231/Domain: link protein repeat homology <LNK1>
 F;222-333/Domain: link protein repeat homology <LNK2>
 F;110, 232, 317/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F;355, 360/Binding site: keratan sulfate (Thr) (covalent) #status experimental

Query Match 8.6%; Score 160; DB 2; Length 370;
 Best Local Similarity 35.8%; Pred. No. 3.7e-06;
 Matches 34; Conservative 16; Mismatches 41; Indels 4; Gaps 2;

Qy 57 VFLHRSPLGQYKLTEDKARACANEATMATYNYNOLSYKOKA 98WLEGRVAYPT 116
 Db 138 VEHYRAIXRTYLDFDRQACLQNSAIIATEPQLOQAVEDGFQCDAGWLADOTVRYPI 197

Qy 117 AFASQNC--GSGVYVGIVDYGPRPKNSEMWDFVCY 148
 Db 198 HTPREGYCDKDEFPGVYGR-DTNEYDVCYCF 231

Search completed: April 4, 2001, 13:04:41
 Job time: 44 sec

F;126, 239, 333, 387, 434, 602, 657, 737, 1898/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;371, 376/Binding site: keratan sulfate (Thr) (covalent) #status predicted

Result No.	Score	Query Match	Length	DB ID	Description
1	262	14.1	277	1	US-08-024-868-2
2	262	14.1	277	2	US-08-024-868-2
3	262	14.1	277	4	PCT-US95-04353-4
4	167	9.0	326	1	US-08-225-477B-3
5	167	9.0	326	4	PCT-US95-04353-6
6	160	8.6	98	2	US-08-242-097-4
7	157.5	8.5	371	4	US-08-225-477B-8
8	157.5	8.5	371	4	PCT-US95-04353-8
9	157.5	8.5	1257	1	US-08-34-128B-49
10	157	8.5	329	4	PCT-US95-04353-3
11	157	8.5	328	1	US-08-225-477B-5
12	156	8.4	333	1	US-08-225-477B-4
13	156	8.4	333	4	PCT-US95-04353-7
14	155.5	8.4	912	4	PCT-US95-04353-8
15	153.5	8.3	97	2	US-08-34-128B-49
16	150.5	8.1	328	4	PCT-US95-04353-5
17	150.5	8.1	328	4	PCT-US95-04353-5
18	150.5	8.1	2409	5	5180808-2
19	148	8.0	908	1	PCT-US95-03147-3
20	147	7.9	378	1	US-08-225-477B-9
21	147	7.9	378	4	PCT-US95-04353-9
22	107	5.8	322	2	US-08-89-280-2
23	105	5.7	810	2	US-08-820-170A-14
24	105	5.7	810	3	US-09-055-699-34
25	105	5.6	810	3	US-09-273-565-34
26	103.5	5.6	717	3	US-08-872-555-9
27	99	5.3	735	3	US-09-191-047-9
28	98.5	5.3	735	1	US-07-946-497-2

GenCore version 4.5	Sequence 7, Appli
Copyright (c) 1993 - 2000 Compugen Ltd.	Sequence 7, Appli
OM protein - protein search, using sw model	Sequence 2, Appli
Run on: April 4, 2001, 13:03:56 ; Search time 13.89 Seconds (without alignments)	Sequence 2, Appli
456.360 Million cell updates/sec	Sequence 2, Appli
Title: US-09-466-778-11	Sequence 2, Appli
Perfect score: 1857	Sequence 2, Appli
Sequence: MTGPKHKCECKSHYVGDL.....ALAAYSYFRINRKTFIGFXHF 353	Sequence 2, Appli
Scoring table: BLOSUM62	Sequence 2, Appli
Gapext: 0.5	Sequence 2, Appli
Searched: 174772 seqs, 17957048 residues	Sequence 2, Appli
Total number of hits satisfying chosen parameters: 174772	ALIGNMENTS
Minimum DB seq length: 0	RESULT 1
Maximum DB seq length: 2000000000	US-08-024-868-2
Post-processing: Minimum Match 08	Sequence 2, Application US/08024868
Maximum Match 100%	Patent No. 5386013
Listing first 45 summaries	GENERAL INFORMATION:
Database : Issued_Patents_AA:*	APPLICANT: Lee, Tae Ho
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*	APPLICANT: Wisniewski, Hans Georg
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*	APPLICANT: Vilcek, Jan
3: /cgn2_6/prodata/2/1aa/6.COMB.pep:*	TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA
4: /cgn2_6/prodata/2/1aa/PCMS_COMB.pep:*	NUMBER OF INVENTION: Coding Therefor and Uses Thereof
5: /cgn2_6/prodata/2/1aa/backfile1.pep:*	NUMBER OF SEQUENCES: 2
Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.	CORRESPONDENCE ADDRESS:
	ADDRESSEE: Browdy and Neimark
	STREET: 419 Seventh Street, NW
	CITY: Washington
	STATE: DC
	COUNTY: US
	ZIP: 20004
	COMPUTER READABLE FORM:
	MEDIUM TYPE: Floppy disk
	COMPUTER: IBM PC Compatible
	OPERATING SYSTEM: PC-DOS/MS-DOS
	SOFTWARE: PatentIn Release #1.24
	CURRENT APPLICATION DATA:
	APPLICATION NUMBER: US/08/024,868
	FILING DATE:
	CLASSIFICATION: 435
	PRIOR APPLICATION DATA:
	APPLICATION NUMBER: US/07/642,312
	FILING DATE:
	ATTORNEY/AGENT INFORMATION:
	NAME: Livnat, Shmuel
	REGISTRATION NUMBER: 33,949
	REFERENCE/DOCKET NUMBER: VILCEK=1
	TELECOMMUNICATION INFORMATION:
	TELEPHONE: 212-628-5197
	TELEFAX: 212-737-3528
	INFORMATION FOR SEQ ID NO: 2:
	SEQUENCE CHARACTERISTICS:
	LENGTH: 277 amino acids
	TYPE: amino acid
	TOPOLOGY: linear
	MOLECULE TYPE: protein
	US-08-024-868-2
	Query Match 52 DTTVGVEHRLSPGLQYKLTDFKARECANEAATMATYQQLSYXOKAYHLC5AGWLETGR 111

Db 32 ERAAGVYHREARSQKYKLYTAEKAVCEFGGHLYTAKIGPHVCAAGWMAGKR 91
 Qy 112 VAYPTAFASNCGSGVGVGIVDGYGRPNKSEMWDFCYRMKDYN C 155
 Db 92 VGYPIVKPGNCGFKGTKGIDYGLRLNRSERWDAYCYNPHAKEC 135

RESULT 2

US-08-242-097-2

; Sequence 2, Application US/08242097
 ; Patent No. 5846763

; GENERAL INFORMATION:

; APPLICANT: Lee, Tae Ho

; ATTORNEY/AGENT: Wiesniewski, Hans Georg

; TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESS: Browdy and Neimark
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.

; ZIP: 20004

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: 13 MAY 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Browdy, Roger L.

; REGISTRATION NUMBER: 25, 618

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-628-5197

; TELEFAX: 212-737-3528

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 277 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; PCT-US96-11995-1

; RESULT 4

US-08-225-477B-6

; Sequence 6, Application US/08225477B
 ; Patent No. 5635370

; GENERAL INFORMATION:

; APPLICANT: Susan Hochfield

; ATTORNEY/AGENT: Diane M. Jaworski

; TITLE OF INVENTION: BEHAB, A Brain Hy-

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESS: St. Onge Steward Johnston & Reens
 ; STREET: 986 Bedford Street
 ; CITY: Stamford

; STATE: CT

; COUNTRY: United States

; ZIP: 06905

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" 1.44 Mb diskette

; COMPUTER: IBM PC

; OPERATING SYSTEM: MS DOS

; SOFTWARE: Word Processor

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/225, 477B

; FILING DATE: April 8, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Mary M. Krinsky

; RESULT 3

PCT-US96-11995-1

; Sequence 1, Application PC/TUS9611995

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: A Pharmaceutical Composition Containing TSG-6

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

REGISTRATION NUMBER: 32423
 REFERENCE/DOCKET NUMBER: 1751-P0004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 203-324-6155
 TELEFAX: 203-327-1096
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 326 residues
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE:
 DESCRIPTION: polypeptide
 FEATURE: Functional domains
 NAME/KEY: rat link protein
 PUBLICATION INFORMATION:
 AUTHORS: Doege, K., Hassell, J.R., Carter, B., and Yamada, Y.
 TITLE: Link protein cDNA sequence reveals a tandemly repeated protein sequence.
 JOURNAL: Proc. Natl. Acad. Sci. USA
 VOLUME: 83
 PAGES: 3761-3765
 DATE: 1986
 S-08-225-477B-6

Query Match 5
 Best Local Similarity 9.08; Score 167; DB 1; Length 326;
 Matches 31.5%; Pred. No. 3.2e-10; Indels 30; Gaps 5;
 保守性 21; Mismatches 49;

Y 5 GKKCKCEKSHYVGDLNCEPEQLPDIRCLQNGOCHA DAKCVDLHQDTIVGFHLRSPL 64
 b 109 GRYKCE---VIEGL-----EDDTAVALEQGV--VVFYFPRL 141
 Y 65 GOYKLTFDKAREACANEATMATHYNOIYXQAKYHICSAQWLETRGVAPTAFASONCG 124
 b 142 GRYNNFHEARQACLDQDAVIASFQDLYDAWGGLDWCNAQWLSQDSVQXPITKPREPCG 201
 Y 125 -SGVGVIVDYGPRPNKSEMWDYFCY 148
 b 202 GQNTVPGVNYG-FWDKDSRYDVFCF 226

Query Match 6
 Best Local Similarity 9.08; Score 167; DB 4; Length 326;
 Matches 31.5%; Pred. No. 3.2e-10; Indels 30; Gaps 5;
 保守性 21; Mismatches 49;

Y 5 GKHGECKSHYVGDLNCEPEQLPDIRCLQNGOCHA DAKCVDLHQDTIVGFHLRSPL 64
 b 109 GRYKCE---VIEGL-----EDDTAVALEQGV--VVFYFPRL 141
 Y 65 GOYKLTFDKAREACANEATMATHYNOIYXQAKYHICSAQWLETRGVAPTAFASONCG 124
 b 142 GRYNNFHEARQACLDQDAVIASFQDLYDAWGGLDWCNAQWLSQDSVQXPITKPREPCG 201
 Y 125 -SGVGVIVDYGPRPNKSEMWDYFCY 148
 b 202 GQNTVPGVNYG-FWDKDSRYDVFCF 226

RESULT 5
 CT-US95-04353-6
 Sequence 6, Application PC/TU9504353
 GENERAL INFORMATION:
 APPLICANT: Susan Hockfield
 APPLICANT: Diane M. Jaworski
 TITLE OF INVENTION: BEHAB, A Brain Hyaluronan-Binding Protein
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: St. Onge, Stewart, Johnston & Reens
 STREET: 986 Bedford Street
 CITY: Stamford
 STATE: CT
 COUNTRY: United States
 ZIP: 06905

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" 1.44 Mb diskette
 COMPUTER: IBM PC
 OPERATING SYSTEM: MS DOS
 SOFTWARE: Word Processor
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04353
 FILING DATE:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 08/225,477
 FILING DATE: April 8, 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Mary M. Krinsky
 REGISTRATION NUMBER: 32423
 REFERENCE/DOCKET NUMBER: 1751-P0004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 203-324-6155
 TELEFAX: 203-327-1096
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 326 residues
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE:
 DESCRIPTION: polypeptide
 FEATURE: Functional domains
 NAME/KEY: rat link protein
 PUBLICATION INFORMATION:
 AUTHORS: Doege, K., Hassell, J.R., Carter, B., and Yamada, Y.
 TITLE: Link protein cDNA sequence reveals a tandemly repeated protein sequence.
 JOURNAL: Proc. Natl. Acad. Sci. USA
 VOLUME: 83
 PAGES: 3761-3765
 DATE: 1986
 S-08-225-477B-6

RESULT 6
 US-08-242-097-4
 Sequence 4, Application US/08242097
 ; Patent No. 5846763
 GENERAL INFORMATION:
 APPLICANT: Lee, Tae Ho
 APPLICANT: Wisniewski, Hans Georg
 APPLICANT: Vilcek, Jan
 APPLICANT: Vilcek, Jan
 TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Browdy and Neimark
 STREET: 419 Seventh Street, N.W., Suite 300
 CITY: Washington
 STATE: D.C.
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patient In Release #1.0
 APPLICATION NUMBER: US/08/242,097
 FILING DATE: 13-MAY-1994
 ATTORNEY/AGENT INFORMATION:
 CLASSIFICATION: 435

NAME: BRONDY, Roger L. ; OTHER INFORMATION: polypeptide encoded by (and set out under) SEQ ID NO 1
 REFERENCE/DOCKET NUMBER: LEE26/VT1CEK-1B ; OTHER INFORMATION: US-08-225-477B-8
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-628-5197
 TELEFAX: 212-731-3528
 SEQUENCE CHARACTERISTICS:
 INFORMATION FOR SEQ ID NO: 4 :
 LENGTH: 98 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-242-097-4

Query Match 8.5%; Score 157.5; DB 1; Length 371;
 Best Local Similarity 27.9%; Pred. No. 4.3e-09;
 Matches 46; Conservative 22; Mismatches 60; Indels 37; Gaps 5;

Qy 10 ECKSHYVGDGGLNCEPEQLPIDRCLOD- ; NCQCHADAKCVDLHF 50
 Db 201 QCDAGWLSDQTVRYLQNPREACYGDMDSYGPVNYGVGPDDLYDVVYCAE --DIN- 255
 Qy 51 QDTTVGVFHLRSPLQGXKLTEDKAREACANEAAATMAYNQLSYXQKAKYHLCAGWLTG 110
 Db 256 ----GELFLGAPPG-KLTWEEARDYCLERGAQIATSGQLYAAWNGGLDRCSPGWLDG 308

Query Match 8.6%; Score 160; DB 2; Length 98;
 Best Local Similarity 30.2%; Pred. No. 3e-10; Matches 34; Conservative 17; Mismatches 41; Indels 2; Gaps 1;

Qy 57 VFHDRSPQSKYKLTEDKAREACANEAAATMAYNQLSXQKAKYHLCAGWLTGRAYPT 116
 Db 3 VFPFPRLGRYNLNFHEARQACLQDAVIASTFQLYDAWRGGSDGWSQYPI 62

RESULT 8
 PCT-US95-04353-8
 Sequence 8, Application PC/TUS9504353
 GENERAL INFORMATION:
 APPLICANT: Susan Hockfield
 APPLICANT: Diane M. Jaworski
 APPLICANT: BEHAB, A Brain Hy-
 TITLE OF INVENTION: Luronan-Binding Protein
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: St. Ongé, Steward Johnston & Reens
 STREET: 986 Bedford Street
 CITY: Stamford
 STATE: CT
 COUNTRY: United States
 ZIP: 06905
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" 1.44 Mb diskette
 COMPUTER: IBM PC
 OPERATING SYSTEM: MS DOS
 SOFTWARE: Word Processor
 CURRENT APPLICATION NUMBER: PCT/US95/04353
 APPLICATION NUMBER: PCT/US95/04353
 FILING DATE:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 08/225,477
 PATENT NUMBER: US5635370
 ATTORNEY/AGENT INFORMATION:
 APPLICANT: Susan Hockfield
 APPLICANT: Diane M. Jaworski
 TITLE OF INVENTION: BEHAB, A Brain Hy-
 NUMBER OF SEQUENCES: 11
 TITLE OF INVENTION: Luronan-Binding Protein
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESS: St. Ongé, Steward Johnston & Reens
 STREET: 986 Bedford Street
 CITY: Stamford
 STATE: CT
 COUNTRY: United States
 ZIP: 06905
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" 1.44 Mb diskette
 COMPUTER: IBM PC
 OPERATING SYSTEM: MS DOS
 SOFTWARE: Word Processor
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/225,477B
 FILING DATE: April 8, 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Mary M. Krinsky
 REGISTRATION NUMBER: 32423
 REFERENCE/DOCKET NUMBER: 1751-P0004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 203-327-1096
 TELEFAX: 203-327-1096
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 371 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 DESCRIPTION: polypeptide
 FRAGMENT TYPE: entire sequence
 IMMEDIATE SOURCE: rat brain
 FEATURE:
 NAME/KEY: rat BEHAB
 OTHER INFORMATION: polypeptide encoded by (and set out under) SEQ ID NO 1
 PCT-US95-04353-8

Query Match 9 Score 157.5; DB 4; Length 371;
 Best Local Similarity 27.9%; Pred. No. 4.3e-09; Indels 37; Gaps 5;
 Matches 46; Conservative 22; Mismatches 60;

Qy 10 ECKSHYVGDLINCEPEQLPDRCLQD-----
 Db 201 QCDAGNLSDOTVRYPIQNPREACTGMDGPYGVRYNQGPDDLYDVCYAE---DLEN- 255

Qy 51 QDTIVGVFHLSPLGQKLFDKAREACANEAAATMAYNQLXKQKXHLSAGWLETG 110
 Db 256 -----GELFLGAPP-KLTWEAARDYCLERGAQFQASTQLYAAWNGGLDRCSPGWLADG 308

Qy 111 RVAYPTAFASQNCGSSGVGVGTVYDGBRPN----KSEMWDVFCYR 149
 Db 309 SVYRPLTPSGRCGGCLPQYKTLFLFPNQTGFPSSKQNRFNVYCFR 353

RESULT 10
 US-08-225-477B-3
 ; Sequence 3, Application US/08225477B
 ; Patent No. 5635270
 ; GENERAL INFORMATION:
 ; APPLICANT: Susan Hockfield
 ; TITLE OF INVENTION: BEHAB, A Brain Hyaluronan-Binding Protein
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: St. Onge, Stewart Johnston & Reens
 ; STREET: 986 Bedford Street
 ; CITY: Stamford
 ; STATE: CT
 ; COUNTRY: United States
 ; ZIP: 06905
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" 1.44 Mb diskette
 ; COMPUTER: IBM PC
 ; OPERATING SYSTEM: MS DOS
 ; SOFTWARE: Word Processor
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/225,477B
 ; FILING DATE: April 8, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maty M. Krinsky
 ; REGISTRATION NUMBER: 32423
 ; REFERENCE/DOCKET NUMBER: 1751-P0004
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 203-324-6155
 ; TELEFAX: 203-327-1096
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 329 residues
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE:
 ; DESCRIPTION: polypeptide
 ; FRAGMENT TYPE: functional domains
 ; FEATURE:
 ; NAME/KEY: rat aggrecan
 ; PUBLICATION INFORMATION:
 ; AUTHORS: Doege, K., Sasaki, M., Horikoshi, K., Bassell, J.R., and Yamada, Y.
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: Margolis-1A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-5528
 ; INFORMATION FOR SEQ ID NO: 49:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1257 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-340-42BB-49

Query Match 9 Score 157; DB 1; Length 329;
 Best Local Similarity 26.8%; Pred. No. 4.1e-09;
 Matches 44; Conservative 23; Mismatches 67; Indels 30; Gaps 6;

Qy 5 GKHKECKSHYVGDLINCEPEQLPDRCLQD-----LODNGOCHADAKCVDLHF 50
 Db 175 GFIHQCD-AGWADQTVYPHTPREGYGDKDEPPGVRITYGIRDNTNEY-DVYC---F 227

Qy 51 QDTIVGVFHLSPLGQKLFDKAREACANEAAATMAYNQLXKQKXHLSAGWLETG 110
 Db 228 AEMEGEFYATSP---EKFTQEAANECTVYARLATTGOLYLAWQQGNDMSAGWLADR 284

Query Match 9 Score 157; DB 1; Length 329;
 Best Local Similarity 34.2%; Pred. No. 2.8e-08; Indels 7; Gaps 3;

Qy 47 DLHQDQTTGVFHLRSPLGQKLFDKAREACANEAAATMAYNQLXKQKXHLSAGW 106
 Db 150 DLVTLTEVGVFHYRARDRVALTAQEECHLSSATIAAPRHLQAAFEDGFNDCAGW 209

Qy 111 RVAYPTAFASQNCGSGVYGI-----VGYPRPNKSEMWDDVFCY 148
 Db 285 SVRYPISKARPNCGGNLGVRVLYHANQTGYPDPSSRYDAICY 328

RESULT 1.1

PCT-US95-04353-3

; Sequence 3, Application PC/TUS9504353
 ; GENERAL INFORMATION:
 ; APPLICANT: Susan Hockfield
 ; APPLICANT: Diane M. Jaworski
 ; TITLE OF INVENTION: BEHAB, A Brain Hyaluronan-Binding Protein
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: St. Ongé, Stewart Johnston & Reens
 ; STREET: 986 Bedford Street
 ; CITY: Stamford
 ; STATE: CT
 ; COUNTRY: United States
 ; ZIP: 06905
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" 1.44 Mb diskette
 ; COMPUTER: IBM PC
 ; OPERATING SYSTEM: MS DOS
 ; SOFTWARE: Word Processor
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/04353
 ; FILING DATE:
 ; APPLICATION NUMBER: 08/225,477
 ; FILING DATE: Apr 11, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mary M. Krinsky
 ; REFERENCE/DOCKET NUMBER: 32423
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 203-324-6155
 ; TELEFAX: 203-327-1096
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 329 residues
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: polypeptide
 ; DESCRIPTION: functional domains
 ; FRAGMENT TYPE: peptide
 ; FEATURE:
 ; NAME/KEY: rat aggrecan
 ; PUBLICATION INFORMATION:
 ; AUTHORS: Doege, K., Sasaki, M., Hori-Gan, E., Hassell, J.R., and Yamada, Y.
 ; TITLE: Complete primary structure of the rat cartilage proteoglycan core protein deduced from CDNA clones.
 ; JOURNAL: J. Biol. Chem.
 ; VOLUME: 262
 ; PAGE: 17757-17767
 ; DATE: 1987
 ; PCT-US95-04353-3

Db 228 AEEMEGEFYATSP---EKETFQEAAANECRTVGRARLATGQQLLAWQGGMDMCSAGNLADR 284
 Qy 111 RVAYPTAFASQNCGSGVYGI-----VGYPRPNKSEMWDDVFCY 148
 Db 285 SVRYPISKARPNCGGNLGVRVLYHANQTGYPDPSSRYDAICY 328

RESULT 1.2

; Sequence 4, Application US/08225477B
 ; Patent No. 5655370
 ; GENERAL INFORMATION:
 ; APPLICANT: Susan Hockfield
 ; APPLICANT: Diane M. Jaworski
 ; TITLE OF INVENTION: BEHAB, A Brain Hyaluronan-Binding Protein
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: St. Ongé, Stewart Johnston & Reens
 ; STREET: 986 Bedford Street
 ; CITY: Stamford
 ; STATE: CT
 ; COUNTRY: United States
 ; ZIP: 06905
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" 1.44 Mb diskette
 ; COMPUTER: IBM PC
 ; OPERATING SYSTEM: MS DOS
 ; SOFTWARE: Word Processor
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/225,477B
 ; FILING DATE: Apr 11, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mary M. Krinsky
 ; REFERENCE/DOCKET NUMBER: 32423
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 203-324-6155
 ; TELEFAX: 203-327-1096
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 333 residues
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: polypeptide
 ; FRAGMENT TYPE: peptide
 ; FEATURE:
 ; NAME/KEY: rat neurocan
 ; PUBLICATION INFORMATION:
 ; AUTHORS: Rauch, U., Karthikeyan, L., Maurel, P., Margolis, R.U., and Margolis, R.K.
 ; AUTHORS:
 ; TITLE: Cloning and primary structure of neurocan, a developmentally regulated, aggregating chondroitin sulfate proteoglycan of brain.
 ; JOURNAL: J. Biol. Chem.
 ; VOLUME: 267
 ; PAGE: 19536-19547
 ; DATE: 1992

US-08-225-477B-4

Query Match 8.5%; Score 157; DB 4; Length 329;
 Best Local Similarity 26.8%; Pred. No. 4.1e-09;
 Matches 44; Conservative 23; Mismatches 67; Indels 30; Gaps 6;

Qy 5 GKHCECKSHYVGQKLNECEPQLPIDRC-----LQDNGQCHADAKCVDLHF 50
 Db 175 GFHQOD--AGWLAQTVRPIHTPRECGYGDKEFPGRTYGRDNTETY-DYCY---F 227
 Qx 51 QDTIVGVFHILRSPLQGQKLTDFKAREACANEATMATYNOLSYXQAKYHLSAGWLETG 110

Query Match 8.4%; Score 156; DB 1; Length 333;
 Best Local Similarity 35.2%; Pred. No. 5.4e-09;
 Matches 37; Conservative 18; Mismatches 46; Indels 4; Gaps 2;

Qy 47 DLHFDQTIVGVFHILRSPLQGQKLTDFKAREACANEATMATYNOLSYXQAKYHLSAGW 106
 Db 126 DLVLTLEVTVFVHYRAARDYALIFAAQEAFLCHSSATIAAPRLNAAEDGFDNCDAGW 185
 Qy 107 LETGRVAYPTAFASQNC---GSGVYGVYDGPRLNEMWDVFCY 148

```

RESULT 13
PCT-US95-04353-4
; Sequence 4, Application PC/TUS9504353
; GENERAL INFORMATION:
; APPLICANT: Susan Hockfield
; APPLICANT: Diane M. Jaworski
; TITLE OF INVENTION: BEHBAB, A Brain Hyaluronic Acid Binding Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: St. Onge Steward Johnston & Reens
; STREET: 986 Bedford Street
; CITY: Stamford
; STATE: CT
; COUNTRY: United States
; ZIP: 06905
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 1.44 Mb diskette
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WORD PROCESSOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04353
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/225,477
; FILING DATE: April 8, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary M. Krinsky
; REGISTRATION NUMBER: 32423
; PREFERENCE/DOCKET NUMBER: 1751-P0004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-324-6155
; TELEFAX: 203-327-1096
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 residues
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: polypeptide
; FRAGMENT TYPE: functional domains
; FEATURE:
; NAME/KEY: rat neurocan
; PUBLICATION INFORMATION:
; AUTHORS: Rauch, U., Karthikeyan, L., and Margolis, R.U., and Margolis, R.K.
; TITLE: Cloning and primary structure of neurocan, a developmentally regulated, aggregating chondroitin sulfate proteoglycan of brain.
; JOURNAL: J. Biol. Chem.
; VOLUME: 267
; PAGES: 19536-19547
; DATE: 1992
PCT-US95-04353-4
Query Match 8.4% Score 156; DB 4; Length 8
Best Local Similarity 35.28; Pred. No. 5.4e-09; Mismatches 46; Inde
Matches 37; Conservative 18; Margolis 46; Inde
Qy 47 DLHQDPTVGPHLSPQKQLTDKARACANEAATMAYNQLSXYQ
Db 126 DLVTLVEVGWVHYAARDVTAPEAQCHLSSATTAAPRHLNAFAF
Qy 107 LETGRVAPTAQSNG--GSGVYIVDVGPRPKNSEWDFVCF 148

```

```

; CITY: Washington
; STATE: D.C.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/242,097
; FILING DATE: 13-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BRONDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: LEE26/VILCEK=1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-628-5197
; TELEFAX: 212-737-3528
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-242-097-5

```

```

Query Match          8.3%  Score 153.5; DB 2; Length 97;
Best Local Similarity 37.9%  Pred. No. 1.5e-09; 3;
Matches 36; Conservative 13; Mismatches 41; Indels 5; Gaps 3;
Qy  57 VFLRLSPPLQYKLTEDKAREACANEATATMAYNQLSTXQKARYHLC5AGWLTGRVAYPT 116
Db   3 VFHXRPGSTRYSLTEERQDACTRTGAISSPRQLQ-AAYAGYEQCDAGWLQDQTVRYP 61
Qy  117 AFASQNC --GSGGVGTVYDGGPRPKSEMMWDYFCY 148
Db   62 VSPRTPCVGDKDSSPGVRYGVPS-SERTDVYCY 95

```

Search completed: April 4, 2001, 13:04:18
 Job time: 22 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 4, 2001, 13:04:01 ; Search time 27.49 Seconds
(without alignments)
1505.071 Million cell updates/sec

Title: US-09-466-778-11
Sequence: 1 MTGPGKHKCECKSHYVGDBGL.....ALAAYSYFRINKRTIGFXHF 353

Scoring table: BLOSUM62

Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.15:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rabbit:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

SPTREMBL.15:*

20 109 5.9 322 4 Q9Y5Y7
21 105 5.7 1584 5 Q91791
22 104 5.6 3507 5 Q23587
23 103.5 5.6 P87357 13 P87357
24 102.5 5.5 778 13 Q91BG4
25 100 5.4 816 11 Q70474
26 99.5 5.4 682 4 Q9TM2
27 99.5 5.4 728 13 Q90656
28 99.5 5.4 802 13 Q57462
29 99.5 5.4 3643 4 Q60494
30 99 5.3 780 11 Q9YX8
31 99 5.3 1440 5 Q20204
32 99 5.3 2906 11 Q9WUH9
33 98 5.3 294 4 Q92493
34 98 5.3 2531 5 Q16004
35 97.5 5.3 780 11 Q08779
36 97.5 5.3 1480 5 Q9V7F8
37 97.5 5.3 1504 5 Q9XY4
38 97.5 5.3 1504 5 Q9V7F9
39 96 5.2 364 11 Q70509
40 96 5.2 742 4 Q9UJ36
41 95.5 5.1 364 6 Q97569
42 95.5 5.1 721 13 Q91902
43 95 5.1 261 10 Q24530
44 94.5 5.1 1328 13 P79754
45 94 5.1 637 10 Q95F50

ALIGNMENTS

RESULT 1
Q9NRY3 PRELIMINARY;
ID Q9NRY3;
AC Q9NRY3;
DT 01-OCT-2000 (TREMBL); 15, Created
DT 01-OCT-2000 (TREMBL); 15, Last sequence update
DT 01-OCT-2000 (TREMBL); 15, Last annotation update
DE C1D4-LIKE PRECURSOR FELT,
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tao Q., Zhang W., Cao X.;
RT "Molecular cloning and characterization of human FELT sharing homology
with Cb4.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFI60476; AAF82298.1; -
SQ SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;

Query Match 94.6%; Score 1757; DB 4; Length 897;
Best Local Similarity 93.8%; Pred. No. 8 6e-162;
Matches 331; Conservative 3; Mismatches 19; Indels 0; Gaps 0;
QY 1 MTGPGKHKCECKSHYVGDBGLNCEPEOLPDIRCLQDNGQCHADAKCVDLHFQDFTVGVHFL 60
Db 489 MTGPGKHKCECKSHYVGDBGLNCEPEOLPDIRCLQDNGQCHADAKCVDLHFQDFTVGVHFL 548
QY 61 RSPLGQYKLTPDKARFACANFAATATYNOISYQOKAKYHLCISGWLIGTRVAYPTAFAS 120
Db 549 RSPLGQYKLTPDKARFACANFAATATYNOISYQOKAKYHLCISGWLIGTRVAYPTAFAS 608
QY 121 QNGSGGVGTVYDGPRLNSKEMWDYFCYRKAQDVKNTXKVYVGDFGFSYSGNLQVLMSPF 180
Db 609 QNGSGGVGTVYDGPRLNSKEMWDYFCYRKAQDVKNTXKVYVGDFGFSYSGNLQVLMSPF 668
QY 161 SITNFTDEVLAYNSNSARGRAGFLEHTDLSIRGTQXPSQNLGENETLSGRDIEHHLAN 240
Db 669 SITNFTDEVLAYNSNSARGRAGFLEHTDLSIRGTQXPSQNLGENETLSGRDIEHHLAN 728

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1757	94.6	897	4 Q9NRY3	Q9nry3 homo sapien
2	1468	79.1	1069	4 Q9UF98	Q9uff98 homo sapien
3	662	35.6	2212	4 Q9Y15	Q9ny15 homo sapien
4	645.5	34.8	2570	11 Q08859	Q08859 mus musculus
5	264	14.2	275	11 Q9787	P79787 gallus gallus
6	177	9.5	2109	13 Q92LX7	Q92LX7 mus musculus
7	169.5	9.1	355	11 Q921X7	Q14594 homo sapien
8	162	8.7	1321	4 Q14594	Q88564 ratmus norvegicus
9	161.5	8.7	1290	13 Q9W6E1	Q9wei gallus gallus
10	161.5	8.7	2394	6 Q77610	Q77610 bos taurus
11	160.5	8.6	192	6 Q02817	Q02817 orctocelus
12	153	8.2	656	6 Q77612	Q77612 bos taurus
13	151.5	8.2	1643	6 Q77611	Q77611 bos taurus
14	151.5	8.2	3391	6 Q9TFB3	Q9tfb3 sus scrofa
15	126.5	6.8	103	6 Q4380	Q4380 orctocelus
16	120.5	6.5	103	6 Q9W6S4	Q9w6s4 gallus gallus
17	110	5.9	396	13 Q9UNF4	Q9unf4 homo sapien
18	109	5.9	302	4 Q9UNF4	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



Query Match 79 '1%; Score 1468; DB 4; Length 1069;
 Best Local Similarity 80.5%; Pred. No. 1.2e-133;
 Matches 284; Conservative 4; Mismatches 29; Indels 36; Gaps 2;

QY 1 VSMFFYNDLVNTTLOTRGSKLLTDOPDPLRPTERCVGDRLTWDICASGNTIHYI 300
 DB 729 VSMFFYNDLVNTTLOTRGSKLLTDOPDPLRPTERCVGDRLTWDICASGNTIHYI 788

QY 301 SRXLKAPPATLXTRXLGIGXXILVNTGAVALAAYSYFRINRTIGFPHF 353
 DB 769 SRXLKAPPATLXTRXLGIGXXILVNTGAVALAAYSYFRINRTIGFPHF 841

RESULT 2
 Q9UEF98 PRELIMINARY; PRT; 1069 AA.
 ID Q9UEF98; PRELIMINARY; PRT; 1069 AA.
 AC Q9UEF98; PRELIMINARY; PRT; 1069 AA.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL 115.7 KDa PROTEIN (FRAGMENT).
 GN DKFZP434E0321.

OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. OC Homo sapiens (Human).
 OC Homo sapiens (Human).
 OC Homo sapiens (Human).
 OX NCBI_Taxid:9606;
 RN [1] 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 RN [1] 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 RN [1] DE HYPOTHETICAL 115.7 KDa PROTEIN (FRAGMENT).
 RN [1] GN DKFZP434E0321.

SEQUENCE FROM N.A.
 TISSUE-TESTIS;
 RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.; Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL133021; CAB61358; 1; -.
 DR HSSP; P98066; 1NSG.
 DR INTERPRO; IPR000538; -.
 DR INTERPRO; IPR000561; -.
 DR INTERPRO; IPR000782; -.
 DR INTERPRO; IPR000791; -.
 DR INTERPRO; IPR000792; -.
 DR INTERPRO; IPR002049; -.
 DR PFAM; PF00008; EGF; 13.
 DR PFAM; PF00193; Xlink; 1.
 DR PRINTS; PRO0011; EGFAMIN.
 DR PROSITE; PRO01265; LINKMODULE.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 2.
 KW Hypothetical protein.
 FT NON_TER 1 1
 SEQUENCE 1069 AA; 115666 MW; 04B0960932164D63 CRC64;

Query Match 35 '6%; Score 662; DB 4; Length 2212;
 Best Local Similarity 40.2%; Pred. No. 4.9e-55;
 Matches 143; Conservative 47; Mismatches 158; Indels 8; Gaps 5;

QY 2 TGPGKHCKECKSHYGDGLNC-EPEQLPTDRCLQDNGOCHADAKCVDLHFDQDTTNGVFLH 60
 DB 1746 TGLINTRICCEIAGYGDGLCQLEEESEPPDRLQGPPCPKEDAMCIDLHFDKRGVFLH 1855

QY 61 RSPLGQYKLTEDKAREACANEAAATMATYNOLSYXQAKYHLSAGILETGVAYTAFAS 120
 DB 1856 QATSGPQGLNFSESEAACEAQAVYLASQFQHLCMGLANGSTAHPPVVFY 1915

QY 121 QRGSGGVGIVDYGGRPKSEMDVFCRMDVNTXKVGIVGDS- YSENLLQVLMF 179
 DB 1916 ADCGNGRGLVYSLGARKNLSEWRDAYCFFQDVACRGRNGFVGDGTSVCKSLQDLYAAT 1975

QY 180 PSLTNFLTEVLAYNSASRRAFELHLDLSIRGLXPQNSGLGENETSGRDLEHHLA 239
 DB 1916 ANFSTYQMLJGYANATQSLQDLEDFDDELTYKTLFVNPNGFVNMTLSGPDLHLAS 2035

QY 181 SLTNFLTEVLAYNSASRRAFELHLDLSIRGLXPQNSGLGENETSGRDLEHHLA 240
 DB 861 CTS-----RTPDDLSIRGLTLPQNSGLGENETSGRDLEHHLA 900

QY 241 VSMFFYNDLVNTTLOTRGSKLLTDOPDPLRPTERCVGDRLTWDICASGNTIHYI 300
 DB 901 VSMFFYNDLVNTTLOTRGSKLLTDOPDPLRPTERCVGDRLTWDICASGNTIHYI 960

QY 301 SRXLKAPPATLXTRXLGIGXXILVNTGAVALAAYSYFRINRTIGFPHF 353
 DB 961 SRXLKAPPATLXTRXLGIGXXILVNTGAVALAAYSYFRINRTIGFPHF 1013





DR WPI: 1999-059056/05
 DR N-PSDB; V71778.
 XX DNA encoding tumour necrosis factor stimulated gene 6 protein -
 PT useful for producing recombinant protein for diagnosis and therapy
 of disease e.g. immune disorders or cancer
 XX
 PS Claim 2; Fig 3A-C; 65pp; English.
 XX This represents a human tumour necrosis factor stimulated gene 6 (TSG-6) protein. A prokaryotic host cell transformed, or a eukaryotic host cell transfected with expression vectors containing the TSG-6 nucleic acid are used to produce recombinant human TSG-6 protein. TSG proteins, peptide fragments of TSG proteins or antibodies to TSG proteins are useful for diagnosis or therapy of diseases mediated by cytokine activity or inactivity, such as immune and autoimmune disorders, infections, inflammatory diseases, neurodegenerative diseases, cancer and alcohol-induced hepatitis.
 CC Sequence 277 AA;
 SQ 14.1%; Score 262; DB 20; Length 277;
 Best Local Similarity 43.3%; Pred. No. 1.6e-19;
 Matches 45; Conservative 16; Mismatches 43; Indels 0; Gaps 0;
 PR 29-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 28-OCT-1997; 97US-0063554.
 PR 29-OCT-1997; 97US-0063435.
 PR 29-OCT-1997; 97US-0063704.
 PR 29-OCT-1997; 97US-0063322.
 PR 29-OCT-1997; 97US-0063338.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063335.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-006248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0065120.
 PR 21-NOV-1997; 97US-0063364.
 PR 24-NOV-1997; 97US-0066742.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0065511.
 XX PA (GETH) GENENTECH INC.
 XX PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 XX DR WPI; 1999-229533/19.
 XX DR N-PSDB; X52252.
 XX PT New isolated human genes and polypeptides used in, e.g. treatment of
 XX gastrointestinal ulceration
 XX PS Claim 12; Fig 78; 320pp; English.
 XX CC Y13344-403 represent secreted and transmembrane human proteins.
 XX The cDNA sequences are obtained from cDNA libraries, prepared from
 XX congenital microvillus atrophy, fetal kidney, fetal liver and fetal retina.
 XX CC cancers such as lung squamous cell carcinoma of the vulva and gliomas,
 XX CC the encoded polypeptides have specific uses based on their homology to
 XX CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 XX CC associated with the preservation and maintenance of gastrointestinal
 XX CC mucosa and the repair of acute and chronic mucosal lesions
 XX CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 XX CC ulceration and congenital microvillus atrophy), skin diseases associated
 XX CC with abnormal keratinocyte differentiation (e.g. psoriasis), epithelial
 XX CC fibromodulin, e.g. for reducing dermal scarring, PRO265 can be used as for
 XX CC as a target for anti-tumor drugs. PRO33 may be used in the treatment
 XX CC of Usher Syndrome or Atrophala areata; PRO69 can be used as an
 XX CC anti-chromotrophic agent; PRO287 polypeptides and portions may have
 XX CC therapeutic applications in wound healing and tissue repair; PRO17 can
 XX CC be used for treating problems of the kidney, uterus, endometrium, blood
 XX CC vessels, or related tissue, e.g. in the heart of genital tract.
 XX

PR 17-OCT-1997; 97US-0062287.
 PR 21-OCT-1997; 97US-0063486.
 PR 24-OCT-1997; 97US-0062814.
 PR 24-OCT-1997; 97US-0062816.
 PR 24-OCT-1997; 97US-0063045.
 PR 24-OCT-1997; 97US-0063120.
 PR 24-OCT-1997; 97US-0063121.
 PR 24-OCT-1997; 97US-0063127.
 PR 24-OCT-1997; 97US-0063128.
 PR 27-OCT-1997; 97US-0063329.
 PR 27-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 29-OCT-1997; 97US-0063554.
 PR 29-OCT-1997; 97US-0063704.
 PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063335.
 PR 31-OCT-1997; 97US-0063870.
 PR 03-NOV-1997; 97US-0064103.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0065120.
 PR 21-NOV-1997; 97US-0063364.
 PR 24-NOV-1997; 97US-0066742.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0065511.
 XX PA (GETH) GENENTECH INC.
 XX PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 XX DR WPI; 1999-229533/19.
 XX DR N-PSDB; X52252.
 XX PT New isolated human genes and polypeptides used in, e.g. treatment of
 XX gastrointestinal ulceration
 XX PS Claim 12; Fig 78; 320pp; English.
 XX CC Y13344-403 represent secreted and transmembrane human proteins.
 XX The cDNA sequences are obtained from cDNA libraries, prepared from
 XX congenital microvillus atrophy, fetal kidney, fetal liver and fetal retina.
 XX CC cancers such as lung squamous cell carcinoma of the vulva and gliomas,
 XX CC the encoded polypeptides have specific uses based on their homology to
 XX CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 XX CC associated with the preservation and maintenance of gastrointestinal
 XX CC mucosa and the repair of acute and chronic mucosal lesions
 XX CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 XX CC ulceration and congenital microvillus atrophy), skin diseases associated
 XX CC with abnormal keratinocyte differentiation (e.g. psoriasis), epithelial
 XX CC fibromodulin, e.g. for reducing dermal scarring, PRO265 can be used as for
 XX CC as a target for anti-tumor drugs. PRO33 may be used in the treatment
 XX CC of Usher Syndrome or Atrophala areata; PRO69 can be used as an
 XX CC anti-chromotrophic agent; PRO287 polypeptides and portions may have
 XX CC therapeutic applications in wound healing and tissue repair; PRO17 can
 XX CC be used for treating problems of the kidney, uterus, endometrium, blood
 XX CC vessels, or related tissue, e.g. in the heart of genital tract.
 XX



PR 14-JAN-1991; 91US-0642312.

XX

PA (DNY) UNIV NEW YORK STATE.

XX

Lee TH, Vilcek J, Wisniewski HG;

XX

WPI; 1992-284330/34.

DR N-PSDA, Q27190.

XX

Tumour necrosis factor-induced (Glyco)protein mol. TSG-6 and its antibody - for treatment and diagnosis of chronic inflammatory conditions e.g. rheumatoid arthritis, infections, sepsis and cancer

PR

PT

TSG-6 DNA

Claim 1; FIG 3; 122PP; English.

PS

Claim 2; Page 58-59; 76PP; English.

XX

XX

XX

CC

PR New TSG-6 compositions to treat inflammatory diseases and cancer - PT can be complexed with inter-alpha-inhibitor, to treat auto-immune PT disease, cancer-related pathology etc. can also be administered as TSG-6 DNA

XX

PS

Claim 2; Page 58-59; 76PP; English.

XX

CC

PR New TSG-6 compositions to treat inflammatory diseases and cancer - PT can be complexed with inter-alpha-inhibitor, to treat auto-immune PT disease, cancer-related pathology etc. can also be administered as TSG-6 DNA

XX

PS

Claim 2; Page 58-59; 76PP; English.

XX

CC

PR 06-FEB-1997.

XX

PF 19-JUL-1996; 96WO-US11995.

XX

PR 20-JUL-1995; 95US-0001311.

XX

PA (DNY) UNIV NEW YORK STATE.

XX

PA Cronstein BN, Vilcek J, Wisniewski H;

XX

DR WPI; 1997-132619/12.

XX

PN

W09704075-A1.

XX

DE 06-FEB-1997.

XX

FT 08-DEC-1998.

XX

PR 13-MAY-1994; 94US-0242097.

XX

PR 13-MAY-1994; 94US-0242097.

PR 14-JAN-1991; 91US-0642312.

PR 01-MAR-1993; 93US-024668.

XX

PA (UINY) UNIV NEW YORK STATE.

XX

PI Lee TH, Vilcek J, Wisniewski H;

XX

PR Lee TH, Vilcek J, Wisniewski H;

XX

FT File Copy





Wed Apr 4 13:03:52 2001

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2001, 13:04:01 ; Search time 27.49 Seconds

(Without alignments)
 1505.071 Million cell updates/sec

Title: US-09-466-778-11

Sequence: 1 MTGPGRHKCECKSHYVGDL.....ALAAYSFERRNRTKIGFXHF 353

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries
 Database : SPTRMBL15:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rabbit:*

12: sp_virus:*

13: sp_vertetebrate:*

14: sp_unclassified:*

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1757	94.6	897	4	Q9NRY3	Q9nry3 homo sapien
2	1468	79.1	1069	4	Q9UF98	Q9uF98 homo sapien
3	662	35.6	2212	4	Q93072	Q93072 homo sapien
4	645.5	34.8	2570	4	Q9N115	Q9n115 homo sapien
5	264	14.2	275	11	Q08859	Q08859 mus musculus
6	177	9.5	2109	13	P79787	P79787 gallus gallus
7	169.5	9.1	355	11	Q921X7	Q921X7 mus musculus
8	162	8.7	1321	4	O14594	O14594 homo sapien
9	161.5	8.7	655	11	Q88564	Q88564 rattus norvegicus
10	161.5	8.7	1290	13	Q9W6E1	Q9w6e1 gallus gallus
11	160.5	8.6	2394	6	O77610	O77610 bos taurus
12	153	8.2	192	6	O02817	O02817 oryctolagus cuniculus
13	153	8.2	656	6	O77612	O77612 bos taurus
14	151.5	8.2	1643	6	O77611	O77611 bos taurus
15	151.5	8.2	3381	6	O77609	O77609 bos taurus
16	126.5	6.8	103	6	Q9TB3	Q9tb3 sus scrofa
17	120.5	6.5	46380	4	O46380	O46380 oryctolagus cuniculus
18	110	5.9	396	13	Q9WS4	Q9ws4 gallus gallus
19	109	5.9	302	4	Q9UNF4	Q9unf4 homo sapien

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Query	Match	Score	1757;	DB	4;	Length	897;
Best Local Similarity	93.8%	Pred. No.	8.6-162;				
Matches	331;	Conservative	3;	Mismatches	19;	Indels	0;
AC	Q9NRY3;	PRELIMINARY;	PRT;	897 AA.			
DT	01-OCT-2000 (TREMBL);	15, Created)					
DT	01-OCT-2000 (TREMBL);	15, Last sequence update)					
DT	01-OCT-2000 (TREMBL);	15, Last annotation update)					
DE	CD4-LIKE PRECURSOR ELL.						
OS	Homo sapiens (Human)						
OC	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Primates; Catarrhini; Hominidae; Homo.						
OX	NCBI_TaxID:9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	Tao Q., Zhang W., Cao X.;						
RT	"Molecular cloning and characterization of human FELL sharing homology with CD44."						
RT	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.						
DR	EMBL: AF160476; AAF8398.1;						
SQ	SEQUENCE 897 AA; 97585 MW; EB920AF36101E388 CRC64;						

Query Match 94.6%; Score 1757; DB 4; Length 897;
 Best Local Similarity 93.8%; Pred. No. 8.6-162;
 Matches 331; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

1 MTGPGRHKCECKSHYVGDL.....ALAAYSFERRNRTKIGFXHF 608
 445.5 34.8 2570 4 Q9N115
 5 008859 11 P79787
 6 177 9.5 2109 13 P79787
 7 169.5 9.1 355 11 Q921X7
 8 162 8.7 1321 4 O14594
 9 161.5 8.7 655 11 Q88564
 10 161.5 8.7 1290 13 Q9W6E1
 11 160.5 8.6 2394 6 O77610
 12 153 8.2 192 6 O02817
 13 153 8.2 656 6 O77612
 14 151.5 8.2 1643 6 O77611
 15 151.5 8.2 3381 6 O77609
 16 126.5 6.8 103 6 Q9TB3
 17 120.5 6.5 46380 4 O46380
 18 110 5.9 396 13 Q9WS4
 19 109 5.9 302 4 Q9UNF4

File Copy



Qy	241	VSMFYFDLYNGTTLQRIGSKLLTIDQDPLHPTETRCYDGRDLEWDICASNGITHYI	300
	729	VSMFYFDLYNGTTLQRIGSKLLTASQDPLQPTETRCYDGRDLEWDICASNGITHYI	788
Qy	301	SRXKAPPAPVTLXKLGXIFXAIITYGAVALLAAYSFRINKRTIGXFH	353
	789	SRPLKAPPAPVTLTNGAGIFFAIIITYGAVALLAAYSFRINKRTIGXFQHF	841
RESULT	2		
RESULT	2		
Q9U98		PRELIMINARY;	
ID	Q9U98	PRT;	1069 AA.
AC			
DT	01-MAY-2000	(TREMBrel. 13, Created)	
DT	01-MAY-2000	(TREMBrel. 13, Last sequence update)	
DT	01-OCT-2000	(TREMBrel. 15, Last annotation update)	
DE		HYPOTHETICAL 115.7 KDA PROTEIN (FRAGMENT).	
GN	DKFP34 E0321.		
OS	Homo sapiens (Human)		
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-TEST;S;		
RA	Blum H, Bauersachs S, Newes H.W., Gassendorfer J., Wiemann S.;		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
DR	AL133021; CAB61358.1; -.		
HSSP	P98066; 1TSG.		
DR	INTERPRO; IPR000538;		
DR	INTERPRO; IPR000561;		
DR	INTERPRO; IPR000782;		
DR	INTERPRO; IPR002049;		
DR	PFAM; PF00008;		
DR	PFAM; PF00193;		
DR	PRINTS; PR00011;		
DR	PRINTS; PR01665;		
DR	LINKMODULE;		
DR	PROSITE; PS00022;		
DR	EGF_1;		
DR	UNKNOWN_2.		
DR	PROSITE; PS01186;		
DR	EGF_2;		
DR	PS01248;		
DR	LAMININ_TYPE_EGF;		
KW	Hypothetical protein.		
FT	NON-TER 1		
SQ	SEQUENCE 1069 AA; 115666 MW;	04B0960932164D63	CRC64;
Query	79.1%	Score 1468; DB 4;	Length 1069;
Best Local Similarity	80.5%	Pred. No. 1..2e-133;	
Matches	284;	Mismatches 4;	Indels 36;
			Gaps
Qy	1	MTGPGRKICKECKSHYDGLANCEPEQPLPDIRCLQDNGQCHADAKCVDLHFQDFTTGVFHL	60
Db	697	MTGPGRKICKECKSHYDGLANCEPEQPLPDIRCLQDNGQCHADAKCVDLHFQDFTTGVFHL	756
Qy	61	RSPLQGYKLTPDKARECANERATMAYTQQLSYXQAKYHLCQSAGMLETGRVAYTAFAS	120
Db	757	RSPLQGYKLTPDKARECANERATMAYTQQLSYXQAKYHLCQSAGMLETGRVAYTAFAS	816
Qy	121	QNCGSGVGVIVDYGPRPKSEMWDFCYRMEDVNTPKVGVYGVGFSYSGNLQLQVMSFP	180
Db	817	QNCGSGVGVIVDYGPRPKSEMWDFCYRMKG-----SAGLFGQLSSRP	860
Qy	181	SITNFFTEVLAGNSSSAARGAFLHITLDSRGLTPXQNGLGENETLSGRDIIBHLAN	240
Db	861	CYS-----RTPDLSRGTLYPQNSGLGENETLSGRDIIBHLAN	900
Qy	241	VSMFYFDLYNGTTLQRIGSKLLTIDQDPLHPTETRCYDGRDLEWDICASNGITHYI	300
Db	901	VSMFYFDLYNGTTLQRIGSKLLTASQDPLQPTETRCYDGRDLEWDICASNGITHYI	960
Qy	301	SRXKAPPAPVTLXKLGXIFXAIITYGAVALLAAYSFRINKRTIGXFH	353
Db	961	SRPLKAPPAPVTLTNGAGIFFAIIITYGAVALLAAYSFRINKRTIGXFQHF	1013

```

RESULT 3
Q93072 PRELIMINARY; PRT; 2212 AA.
ID Q93072;
AC DT 01-FEB-1997 [TREMBLrel. 02, Created]
DT 01-FEB-1997 [TREMBLrel. 02, Last sequence update]
DT 01-OCT-2000 [TREMBLrel. 15, Last annotation update]
DE MYELOBLAST KIAA0246 PROTEIN (FRAGMENT).
DE KIAA0246.
GN OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID:9606;
RN [1]
RN RP SEQUENCE FROM N.A.
RN TISSUE=BONE MARROW;
RX MEDLINE="9719544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyaura N., Nomura N.;".
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-324(1996).
DR EMBL; DB/433; BAA337.1; -.
DR HSSP; PR0066; 1TSG;
DR INTERPRO; IPR000538; -
DR INTERPRO; IPR00531; -
DR INTERPRO; IPR000732; -
DR INTERPRO; IPR001128; -
DR INTERPRO; IPR002049; -
DR PFAM; PF00008; EGF; 13.
DR PFAM; PF00193; Xlink; 1.
DR PRINTS; PR01265; LINKMODULE.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 13.
DR PROSITE; PS01241; LINK; UNKNOWN_1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 2.
KN Glycoprotein.
FT NON_TER 1
SO SEQUENCE 2212 AA; 237451 MW; 4A95460504129134 CRC64;

```

Query, Match	Score 1468; DB 4; Length 1069;	Score 662; DB 4; Length 2212;
Best Local Similarity	Pred. No. 1.2e-133;	Pred. No. 4.9e-15;
Matches	Mismatches 4;	Mismatches 158;
Qy 1 MTGPGXHKCECKSHVYGDGLNCEPPQLPDIRCLQDGQCHADAKVCDLHFQDTTGYFHL 60	Qy 2 TGPGRHKCBCKSHVYGDGLNCEPEQLPDIRCLQDGQCHADAKVCDLHFQDTTGYFHL 60	
Db 697 MTGPGXHKCECKSHVYGDGLNCEPPQLPDIRCLQDGQCHADAKVCDLHFQDTTGYFHL 756	Db 1796 TGLTRRCCHAYVGDQCLCSEEPVDRCLQGPCHSDAMCTDLHQEKTAG	
Qy 61 RSPLGQKLTEDPKAREACANEAAATMATYQNLSTXQAKVHLC5AGWLETGRYAPFAFAS 120	Qy 61 RSPLGQKLTEDPKAREACANEAAATMATYQNLSTXQAKVHLC5AGWLETGRYAPFAFAS 120	
Db 757 RSPLGQKLTEDPKAREACANEAAATMATYQNLSTXQAKVHLC5AGWLETGRYAPFAFAS 816	Db 1856 QATSPGGINFSRERAEEAAGQAVLAQFPQLSAAGQFLGFLCLMRLANSLATPVY	
Qy 121 QNCGSQVQIVDYGPRPKNSEMDVFCYRMQDNCYVQGDFGSVSGNLQVMSFP 180	Qy 121 QNCGSQVQIVDYGPRPKNSEMDVFCYRMQDNCYVQGDFGSVSGNLQVMSFP 180	
Db 817 QNCGSQVQIVDYGPRPKNSEMDVFCYRMQDNCYVQGDFGSVSGNLQVMSFP 860	Db 1916 ADCNGRGRYIVSGLARKULSERWDACFRYDVA�CRNGFVGDIISTICCKKLDD	
Qy 181 SLTNFLTEVLAYNSSSARGRAFLEHTDLSRGTFEXPONSGLGENETLSQRDIEHHLAN 240	Qy 180 PSLNFLTEVLAYNSSSARGRAFLEHTDLSRGTFEXPONSGLGENETLSQRDIEHHLAN 240	
Db 861 CTS-----RTPDDLSIRGLFVPPONGENELSGRDIEHHLAN 900	Db 1976 ANFSTEGMULGANTORGULDFDDELYKTFVNEGEVDMNTLSQPLP	
Qy 241 VSMFFYNDLYNGTTLQTRLGSKLLITDRODPLHPTETCYDGRDTLENDICASNGITHVI 300	Qy 240 NVSMFFYNDLYNGTTLQTRLGSKLLITDRODPLHPTETCYDGRDTLENDICASNGITHVI 300	
Db 901 VSMFFYNDLYNGTTLQTRLGSKLLITASQDPLQFETREVGDRAILQDIFASNGITHVI 960	Db 2036 NATLISAN-ASQGKLPLAHSGSLSLISDAGPDNSSAPVAPGTVVYSRIVWDIMA	
Qy 301 SRLXKAPAPVTLXITXLGIFXXIIIITGAVALAAYASYERIRNKTIGFXIF 353	Qy 297 THVIRSLXKAPAPVTLXITXLGIFXXIIIITGAVALAAYASYERIRNKTIGFXIF 353	
Db 961 SRLXKAPAPVTLXITXLGIFXXIIIITGAVALAAYASYERIRNKTIGFXIF 1013	Db 2095 THALASPLAPOPVALAPEAPPVAGVLAAGALLGLVAGAYLXRANGKPMF	

